SAPIENZA
UNIVERSITÀ DI ROMA

UNIVERSITÀ DI ROMA "LA SAPIENZA"
Dipartimento di Matematica

STOCHASTIC MODELS OF BIOLOGICAL EVOLUTION

Davide Palmigiani

Dottorato di Ricerca in Matematica, XXXI Ciclo

Advisor:
Prof. Roberto Natalini
IAC - CNR

I would like to thank my advisor, Roberto Natalini: thanks to his support, he guided me (not only) in this research.

Thanks to the coauthors of my works Anna Lisa Amadori and Giovanni Sebastiani, and to Giovanni Luca Torrisi, for helpful discussions.

I thank my two referees too, Christoph Hauert and Luigi Preziosi, for their precious and timely advice.

Contents

Introduction

Charles Darwin was a keen observer. He analyzed animal breeders, their methods, their results. In his masterpiece *On the Origin of Species* [17], he recognized that Nature was actually a gigantic breeder, giving a definition of the concept of *natural selection.* Indeed, Charles Darwin was not a great mathematician: "I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics, for men thus endowed seem to have an extra sense [18]". There are no equations in *the Origin of Species.* The one problem that the naturalist did not solve concerned the mechanism that could maintain diversity in a population for natural selection to operate. It is interesting to note how the answer was present in the same years, as Gregor Mendel, Austrian monk and botanist, was making his experiments on plant inheritance, using statistical techniques to confirm his results. Mendel's work had already been published [63], but not in English, and remained for years ignored in the Annals of the Brno Academy of Sciences.

Only in the 1900s a simple mathematical equation was formulated, today recognized as one of the fundamental principles of evolution under sexual reproduction, the Hardy-Weinberg law [42] (independently found by the British G.H. Hardy and the German physician W. Weinberg): Mendelian inheritance does lead to maintenance of genetic diversity under random mating. Mendelian genetics and Darwinian evolution were unified in the new discipline of *mathematical biology* in the early twentieth century, when Fisher, Haldane and Wright produced the modern evolutionary synthesis [27, 39, 94]. Through their work, fundamental concepts of evolution, selection, and mutation were for the first time embedded in a precise mathematical framework.

This thesis, on the footsteps of these pioneers, focus on population dynamics, the branch of mathematical biology that studies the composition of populations as dynamical systems and the biological and environmental processes driving them. We present some classical deterministic models of population dynamics and introduce the selection equation (replicator dynamics) and the quasi-species equation for mutations (see Chapter 1), that presuppose mutations occurring at a homogeneous rate. After that, we recall concepts of Probability theory (in Chapters 2 and 3) as point processes and stochastic integrals, which will be used in the construction of the models in Chapters 4, 5 and 6. Those models start from the assumption that mutations must be modeled as rare events that, through jumps, can change the frequencies in the population. We show in this introduction the fundamental steps.

The first principle of population dynamics is widely regarded as the EXPONENTIAL LAW OF MALTHUS proposed in the book *An Essay on the Principle of Population* [55], where the author predicts a difficult future for humanity, assuming that there is an exponential growth of individuals, but linear growth of food resources. This book was an inspiration for Darwin, who brought it with him when he embarked on the *Beagle.* The variation in time of the number of individuals of a population indicated with $x(t)$ is linked to a growth rate R, obtaining the

simple equation $\dot{x} = Rx$.

The initial exponential growth of a population sooner or later slows down, and we can consider the rate R as a linear function of the population x, i.e. $k(1-x)$. This choice leads to the LOGISTIC EQUATION (P.F. Verhulst [92]):

$$\begin{cases} \dot{x} = kx(1-x), \\ x(0) = x_0 \in [0,1]. \end{cases}$$

The next step is to assume that more than one species can interact with each other, as a dynamic between prey and predator, as done by Alfred Lotka and Vito Volterra [54]:

$$\begin{cases} \dot{x} = (\alpha - \beta y)\, x, \\ \dot{y} = (-\gamma + \delta x)\, y, \\ x(0) = x_0 \in \mathbb{R}^+, \\ y(0) = y_0 \in \mathbb{R}^+. \end{cases}$$

where the number of the prey and the predator is indicated respectively by the functions $x(t)$ and $y(t)$. The parameters α and γ takes into account the Malthusian growth of prey and predator, β and δ are the interaction terms.

A step further in the development of population dynamics for Evolution occurs when Taylor and Jonker [88] introduced the REPLICATOR DYNAMICS, a system that describes the evolution of the frequencies of individuals interacting with each other. Assume a population divided into D types with frequencies $x_1, \ldots, x_D$, we have then a dynamic for $x(t)$ on the symplex $\mathcal{S}^D$. The rate of increase of the i-th type $\dot{x}/x$ is equal to the difference between the reproductive success of an individuals, i.e. the fitness $\Phi_i(x)$, and the average fitness of the population,

$$\bar{\Phi}(x) = \sum x_i \Phi_i(x).$$

These choices lead to the *replicator equation*, that can model the essence of selection and adaptation:

$$\begin{cases} \dot{x}_i = x_i \left(\Phi_i(x) - \bar{\Phi}(x) \right), \\ i = 1, \ldots, D. \end{cases} \tag{1}$$

The path of the replicator dynamics crosses in the second half of 1900 that of a branch of mathematics just born and that initially did not seem to have points of contact with Biology, i.e. *Game theory*.

Game Theory is a mathematical theory born in the first half of 1900s that, as their creators John von Neumann and Oskar Morgenstern intended [71, 67] is able to study situations, called *games,* where behavior of *players* (that can be individuals, societies) in strategic and economic decisions is considered. Game theory was further developed by John Nash [69]. At the basis of this theory is the concept of *rational decision-maker*, i.e. individuals that are able to make decision that maximize an expected advantage, or utility. Take as example the PRISONER'S DILEMMA where two criminals are accused of committing a crime and, after the arrest, are divided in two different cells, without the possibility of communicating; two choices are given: *betray* by testifying that the other committed the crime, or *remain silent*. Each pair of choices made by the criminals provides a certain gain, a greater or lower number of years in jail. The payoff can be represented using a matrix $\mathcal{U}$:

$$\mathcal{U} = \begin{pmatrix} -4 & 0 \\ -5 & -2 \end{pmatrix},$$

where the rows represent the choices of the first player (*betray*, *remain silent*), the columns the choices of the second, while the values are the gain of the first player, based on the strategies chosen by both. Choosing the best strategy to adopt according to the different payoffs is the subject of game theory. Interactions of this kind occur in various fields of knowledge such as Economics, Social Sciences and Philosophy. It was John Maynard Smith that successfully applied the theory of games to Biology. As he himself says [57], "paradoxically, Game Theory is better applied to Biology than to economic behavior, to study which it had been invented". According to Maynard Smith the concept of *fitness* and adaptability fits perfectly with the *utility function* of Game Theory, and there are many examples in which biological behaviors acquire a deeper sense in the light of this theory. Take the behavior of some armed species, such as stags, that in the face of a conflict for food, territory, or a companion, engage in a fight that rarely results in a physical confrontation, reducing often to a series of conventional rituals. This behavior can be explained by a game, called *Hawks and Doves*. Suppose there are only two possible behaviors in the population: the first type of stags is aggressive (*Hawks*) and will look for the physical fight whenever possible, they will fight until exhaustion if necessary. The second type of stags is peaceful (*Doves*) and will tend to avoid the struggles, face to face with another peaceful they begin a simulated ritual battle until one goes away, face to face with an aggressive always run away [44]. This game, as the one presented above, can be represented by a payoff matrix:

$$\mathcal{U} = \begin{pmatrix} \frac{G-C}{2} & G \\ 0 & \frac{G}{2} \end{pmatrix}, \tag{2}$$

where $G > 0$ is the gain after a fight, $C > G$ is the cost a Hawk pays after losing a physical fight. When we combine the Replicator Dynamics (1) with GAME THEORY, writing the fitness as linear functions of an appropriate payoff matrix,

$$\Phi(x) = \mathcal{U}x,\ \bar{\Phi}(x) = x \cdot \mathcal{U}x,$$

we can dynamically explain what is about to happen: in a population composed almost entirely of Doves, a few Hawks would increase in number because, meeting only peaceful animals, have a gain almost equal to G. In a population of aggressive Hawks, on the other hand, it would be the Doves to win, because their fitness would remain the same while the Hawks would lose fighting with each other. The system leads the population to converge towards the point of equilibrium with a fraction of Doves equal to $\frac{G}{C}$. The solution of the equations is often given by a dynamic that converge to some stable equilibria, that are *evolutionary stable*, i.e. formed by a population that is resistant to the occasional appearance of invasive minorities.

One of the limitation of the replicator dynamics is the fact that it does not take into account mutations. During replication of a genome, mistakes can happen, and those mistakes are one of the driving forces of evolution. Then, we deal with models that involve mutations, starting from the classical QUASISPECIES DYNAMICS, proposed by Manfred Eigen and Peter Schuster [85]:

$$\begin{cases} \dot{x}_k = \left(\Phi_k - \bar{\Phi}\right) x_k + \sum_{i=1}^{D} \Phi_i m_{ik} x_i, \\ k = 1, \ldots, D. \end{cases} \tag{3}$$

This equation is a variation of the replicator dynamics, with added terms given by the matrix $M = (m_{ik})_{i,k=1,\dots,D}$, effective mutations matrix. The new terms describe the effect of mutations on the dynamics: individuals of type i can mutate and turn into individuals of type k, according to a rate $\Phi_i m_{ik} \geq 0$; on the other hand, the frequency of a given type k can decrease when individuals of that type mutate, $\Phi_k m_{kk} \leq 0$.

In this thesis, we want to highlight how mutations occur at a different time scale than selection, in rare and random moments. For this reason, most of the models presented explain the biological mutation processes using stochastic terms. In particular, *point process* are used to model the mutation jumps, so, to introduce them, the definition of Poisson random measure is necessary. We provide here a construction for such a measure, so as to immediately give it an intuitive interpretation:

Fact. *[50] Let (E, ν) be a measurable space, with ν a σ-finite measure.*

- *Suppose $\nu(E) < +\infty$. Define $\lambda := \nu(E)$ and let $(T_n)_{n\in\mathbb{N}}$ be the jump times of a homogeneous Poisson process on $\mathbb{R}^+$, with intensity λ. Let $(\xi_n)_{n\in\mathbb{N}}$ be independent random variables, identically distributed, uniform on (E, ν). Then, a Poisson random measure $\mathcal{N}(ds, d\xi)$ on $\mathbb{R}^+ \times E$ can be expressed as*

$$\mathcal{N}(ds, d\xi) = \sum_{n\in\mathbb{N}} \delta_{T_n}(ds)\delta_{\xi_n}(d\xi); \tag{4}$$

When the measure $\nu(E)$ is finite, we can think a Poisson random measure as a sequence of random points (T_n, ξ_n) in the space $\mathbb{R}^+ \times E$. The values in $\mathbb{R}^+$ are occurrences of a homogeneous Poisson that has intensity $\nu(E)$. For each event T_n there is a value ξ_n, uniformly sampled in E, with measure ν (see Figure 1 for a visual representation of the construction of $\mathcal{N}$).

A stochastic differential equation with jump terms can be therefore formulated in the following form:

$$X_t = X_0 + \int_0^t b(X_s, s)ds + \int_0^t \int_E K\left(X_{s^-}, \xi\right)\mathcal{N}(ds, d\xi).$$

and we can write:

$$\int_0^t \int_E K(X_{s^-}, \xi)\mathcal{N}\left(ds, d\xi\right) = \sum_{n:\, T_n \leq t} K(X_{T_n^-}, \xi_n);$$

a point process, therefore, evolves deterministically according to the function b, up to the first time T_1. At that moment a jump occurs, and the function K is evaluated in the point (T_1, ξ_1). After that moment, the system continues to move according to the deterministic term, until the subsequent time T_2, and so on.

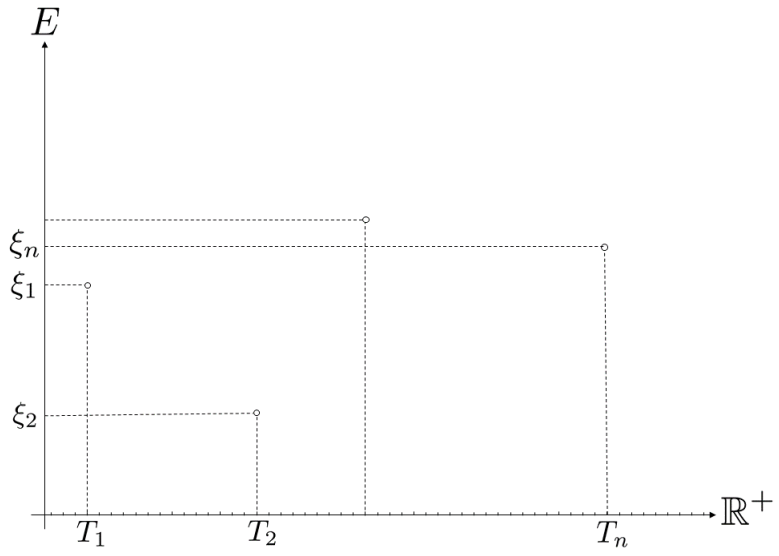


Figure 1: Visual representation of the construction of a Poisson random measure $\mathcal{N}$ on $\mathbb{R}^+ \times E$ (with $\nu(E) < \infty$). The jump times T_n, $n \in \mathbb{N}$ are from a homogeneous Poisson process; the succession of independent marks ξ_n, $n \in \mathbb{N}$ are uniform on E. The function K acts on the space $\mathbb{R}^+ \times E$, i.e. on the pairs (T_n, ξ_n), $n \in \mathbb{N}$.

Another classic example of a random process is the one studied by Robert Brown in 1827, i.e. *Brownian motion*. Once properly formalized, it is possible to include it in the theory of differential equations. Brownian motion is described by the Wiener process W_t, a stochastic process, continuous almost surely, with independent increments distributed as Gaussian random variables. Starting from the Wiener process, it is possible to define the concept of diffusive stochastic differential equation:

$$X_t = X_0 + \int_0^t b(X_s, s)ds + \int_0^t \sigma(X_s, t)dW_s,$$

for $0 \leq t \leq T$. The drift term b represents the deterministic motion of the system, while the diffusion coefficient σ models the random perturbations due to a Brownian motion. As mentioned, jump processes will be used to model mutations, while Brownian motion will model the physical movement of individuals in space.

In the subsequent models we are interested in describing the solutions of stochastic differential equations in terms of deterministic quantities, such as expected values or probability density. It is possible to obtain this results using particular partial differential equations, the Kolmogorov equations (backward and forward). The connection between S.D.E.s and these P.D.E.s is the *infinitesimal generator* of a stochastic process, a partial differential operator that encodes information about the process. The infinitesimal generator of the process X_t is the differential operator $\mathcal{A}$, defined on suitable functions $u(x)$ by:

$$\mathcal{A}u(x) := \lim_{t \to 0} \frac{\mathbb{E}\left(u(X_t)\right) - u(x)}{t}.$$

Fact. *Let X_t the solution of a stochastic differential equation with diffusive and jump terms*

on $\mathbb{R}^n$*:*

$$X_t = X_0 + \int_0^t b(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s + \int_0^t \int_E K\left(X_{s^-}, \xi\right)\mathcal{N}(ds, d\xi).$$

Then the generator $\mathcal{A}$ *takes the form* $\mathcal{L} + \mathcal{I}$*, where:*

$$\mathcal{L}u(x) = \sum_{i=1}^{n} u_{x_i}(x)b_i(x) + \frac{1}{2}\sum_{i,j} u_{x_i x_j}(x)\left[\sigma\sigma^t\right]_{ij}(x),$$

$$\mathcal{I}u(x) = \int_E \left[u(x + K(x, \xi)) - u(x)\right]\nu(d\xi).$$

The generator is used in the Kolmogorov backward or Feynman-Kac equation (which describes the evolution of the expected value of the a process) and its L^2 adjoint is used in the Kolmogorov forward or Fokker–Planck equation (which describes the evolution of the probability density functions of the process). The expected value of a process X at time t, starting from the initial deterministic value x,

$$u(x, t) = \mathbb{E}\left(h(X(t))\mathrm{e}^{-\int_0^t \Lambda(X(u))du} | X(0) = x\right),$$

can be represented as the solution $u(x, t)$ of the Feynman-Kac equation

$$\begin{cases} u_t(x, t) = \mathcal{A}u(x, t) - \Lambda(x)u(x, t), & t > 0 \\ u(x, 0) = h(x), \end{cases}$$

when h is the identity function and $\Lambda = 0$. Assuming that the process X_t has density $\varrho(x, t)$, ϱ is a solution of the equation

$$\begin{cases} \varrho_t = \mathcal{A}^*\varrho, & \text{for } (x, t) \in \mathbb{R}^d \times (0, \infty) \\ \varrho(t = 0) = \varrho_0, & \text{for } x \in \mathbb{R}^d \end{cases} \tag{5}$$

where $\mathcal{A}^*$ is the adjoint in L^2 of the generator of the process, and $\varrho_0(x)$ the density of X_0.

The model in Chapter 4, presented in [4, 3], is a generalization of the replicator mutator model (3) and is a starting point for the two successive models, in Chapter 5 and 6. The main idea is to transform the deterministic term of mutation into a stochastic jump process for the variable X_t, which in rare and random moments brings the type i to mutate and transform into the type j:

$$\begin{cases} X_t^k = X_0 + \int_0^t a_k(X_s)dt + \sum_{i \neq k} \int_0^t \gamma_{ik} X_{s^-}^i dN_s^{ik} - \sum_{i \neq k} \int_0^t \gamma_{ki} X_{s^-}^k dN_s^{ki}. \\ k = 1, \ldots, D \end{cases} \tag{6}$$

The functions $a_k(X) = \left(\Phi_k(X) - \bar{\Phi}(X)\right)X^k$ are the term of the replicator dynamics with $k = 1, \ldots, D$. The processes N_t^{ik} model a mutation with a fixed ancestor i and a descendant of a single different type k, with intensity that depends on the genetic distance between the type i and the type k but also from the selection: the higher is the fitness $\Phi_i(x_{t^-})$, the more

types i will reproduce, more often the offspring will suffer mutations ([4]). For this reason the processes have non homogeneous stochastic intensity $\lambda_{ik}\Phi_i(X_t^-)$. The proportion between the descendants of individuals of type i showing the type k after a mutation is constant and equal to $\gamma_{ik} \in (0, 1]$. Intuitively, the population evolves, starting from time 0, in a deterministic way, according to the law given by a, until the first mutation takes place, i.e. the first jump occurs, at time t_1. At each mutation event, a random variable selects which of the D genomes mutates, and in which of the remaining $D-1$ it transforms. The probability that the jump from i to j is chosen proportional to the quantity $\lambda_{ij}\Phi_i\left(X_{t_1^-}\right)$. After the selection of the pair (i, j), the actual mutation occurs and the proportion $\gamma_{ij}x_i$ of individuals of type i is transformed into individuals of type j. From this point the dynamic is deterministic, with altered proportions, until the following jump.

Starting from model (6), we can obtain Feynman-Kac equations for the expected values:

$$\begin{cases} \partial_t u_k(x,t) = a(x) \cdot \nabla u_k(x,t) + \mathcal{I}u_k(x,t), & x \in \mathcal{S}^n,\ t > 0 \\ u_k(x,0) = x_k, & x \in \mathcal{S}^n, \end{cases}$$

with the generator of the jump process $\mathcal{I}$:

$$\mathcal{I}u(x,t) = \sum_{i \neq j} \lambda_{ij}\Phi_i(x) \left[u(x + \gamma_{ij}x_i(e_j - e_i), t) - u(x,t)\right].$$

In the simplest case, with two species and constant fitness, we can reduce the number of variables, $x_0 = 1 - x$, $x_1 = x$, and write

$$\begin{cases} \partial_t u + (\Phi_0 - \Phi_1)(1-x)x\partial_x u = \lambda_0 f_0 \mathcal{I}_0 u + \lambda_1 \mathcal{I}_1 u, & x \in [0,1],\ t > 0 \\ u(x,0) = x, & x \in [0,1] \end{cases} \tag{7}$$

where

$$\mathcal{I}_0 u(x,t) = u(x + \gamma_0(1-x), t) - u(x,t),$$
$$\mathcal{I}_1 u(x,t) = u(x - \gamma_1 x, t) - u(x,t).$$

The function u represents the expected value $\mathbb{E}\left(x(t) | x(0) = x\right)$. We can compare (7) to an equation of homogeneous transport linked to the replicator mutator (3) and understand that the expected value of the population quantity of the stochastic equation is greater than or equal to that of the deterministic case, therefore rare mutations increase the survival opportunities of the lower-fitness species. The next step is to consider an extension of this model, that takes into account a factor not sufficiently highlighted in previous models, the presence of a spatial heterogeneous environment, in which the population can move.

In Chapter 5 we expand the stochastic model presented in [4] to take into account how the natural environment can modify the interactions between individuals, changing the fitness functions; results presented have been collected in article [5], published in 2017. To introduce heterogeneous environment we increase the observed variables so that the status of the population is described by a pair (x, y): as before $x \in \mathcal{S}^D$ stands for the composition of the population, each x_i being the fraction of individuals of a fixed type, while the new variable $y \in \mathbb{R}^N$ stands for the *position* of the population in the physical N-dimensional space, or can be seen as an external parameter that can change fitness values over time.

Example. Take the two strategy game Hawks vs Doves ($D = 2$), with the payoff matrix $\mathcal{U}$ as in (2). In this new model, we assume that the cost for fighting can depend of y, $C = C(y)$, with the function C designed so that at $y < 0$ the cost lowers and the environment becomes favorable to the Hawks. Otherwise if $y > 0$ environment is more favorable to Doves, with the cost of the fight that increases.

The population x evolves according to replicator dynamics with rare mutations (6):

$$x_t = x_0 + \int_0^t a\,(x_s, y_s)\,ds + \int_0^t \int_E K(x_{s^-}, y_s, \xi)\mathcal{N}(ds, d\xi),$$

where the function $a \in \mathbb{R}^D$ has the fitness allowed to depend from the position y, so that

$$a_k(x, y) = x_k(\Phi(x, y) - \bar{\Phi}(x, y)) \quad \text{as } k = 1, \dots D.$$

Example. The fitness functions for Hawks (x_1) and Doves (x_2), are respectively

$$\Phi_1 = (G - C(y))x_1/2 + Gx_2, \quad \Phi_2 = Gx_2/2,$$

then the replicator dynamics (reducing the coordinates only to $x \in [0, 1]$, fraction of Hawks) is

$$\dot{x} = x\,(1 - x)\,(\Phi_1 - \Phi_2) = x\,(1 - x)\,(G - C(y)x)\,/2.$$

The environmental variable y changes according to a diffusion with drift:

$$y_t = y_0 + \int_0^t v\,(x_s, y_s)\,ds + \int_0^t \sigma\,(x_s, y_s)\,dW_s,$$

where $v \in \mathbb{R}^N$ stands for the *velocity* of the population, σ is an $N \times N$ matrix and W_s is an N-dimensional Brownian motion, describing the random component of the movement. Notice that both the drift and the diffusion may depend on the frequency vector x, allowing retro-actions of population on the environment itself.

Example. In the Hawk vs Doves game, we can take the deterministic velocity field as $v = v(x)$, assuming that, when the concentration of Doves is high ($x \simeq 0$) the population tends to move towards the positive y, favorable for the majority, towards the negative y when $x \simeq 1$.

In general we are facing the following model:

$$\begin{cases} x_t = x_0 + \displaystyle\int_0^t a\,(x_s, y_s)\,ds + \int_0^t \int_E K(x_{s^-}, y_s, \xi)\mathcal{N}(ds, d\xi), \\ y_t = y_0 + \displaystyle\int_0^t v\,(x_s, y_s)\,ds + \int_0^t \sigma\,(x_s, y_s)\,dW_s, \end{cases} \tag{8}$$

of which we are interested in studying the macroscopic function $\varrho(x, y, t) \in [0, 1]$, measuring the probability of finding a population distribution $x \in \mathcal{S}^D$ in the position $y \in \mathbb{R}^N$ at time t. For instance, we may be interested in understanding, at time $t > 0$ the probability of having a high proportion of individuals of a certain type i, or the probability of finding a high proportion of individuals of type i near at the origin.

To obtain a law that describes the evolution of the function ϱ, we calculate initially, in line with the arguments of the model (6), the Feynman-Kac equation of the process, then we

compute $\mathcal{L}^*$, the dual operator in L^2 of the infinitesimal generator. It turns out that, if $\varrho_0(x,y)$ is the probability density of the random pair (x_0,y_0) describing the initial distribution of the population, and if the solution (x_t,y_t) to (8) has a sufficiently smooth probability density $\varrho(x,y,t)$ for $t>0$, then it solves the Fokker-Plank integro-differential equation:

$$\begin{cases} \partial_t \varrho - \frac{1}{2}\sum_{h,k=1}^{N} \partial^2_{y_h y_k}\left((\sigma\sigma^t)_{hk}\varrho\right) + \mathrm{div}_x\left(\varrho a\right) + \mathrm{div}_y\left(\varrho v\right) = \sum_{i=1}^{d} \mathcal{J}_i^*(\Phi_i \varrho) \\ \varrho(x,y,0) = \varrho_0(x,y), \end{cases} \tag{9}$$

in the closed set $(x,y)\in \mathcal{S}^D\times\mathbb{R}^N$ and $t>0$. However, there is no reason to expect that the density function is smooth enough, due to the point process modeling mutation. We therefore choose to write the Fokker-Plank equation formally and then to settle it in the framework of viscosity solution theory. This approach has the advantage of asking very few a-priori regularity and producing well-posed solutions even in this degenerate elliptic, integro-differential case.

A further problem is with operators $\mathcal{J}_i^*$, which are not continuous with respect to x. We therefore switch to another problem which is set into all $\mathbb{R}^D\times\mathbb{R}^N$ and is continuous, extending properly the fitness functions Φ_i, the drift v, the diffusion σ in a bounded smooth way to all $\mathbb{R}^D\times\mathbb{R}^N$, and Concerning the initial datum ϱ_0, it can be extended as $\varrho_0\equiv 0$ outside $\mathcal{S}^D\times\mathbb{R}^N$. Then we solve the following problem:

$$\begin{cases} \partial_t \varrho - \frac{1}{2}\sum_{h,k=1}^{N} \partial^2_{y_h y_k}\left((\sigma\sigma^t)_{hk}\varrho\right) + \mathrm{div}_x\left(\varrho a\right) + \mathrm{div}_y\left(\varrho v\right) + c\varrho = \sum_{i=1}^{d} \mathcal{J}_i^*(\Phi_i \varrho) \\ \varrho(x,y,0) = \varrho_0(x,y), \end{cases} \tag{10}$$

for $(x,y)\in\mathbb{R}^D\times\mathbb{R}^N$ and $t>0$. In view of these remarks, we define a solution of the problem (9) in the symplex $\mathcal{S}^D$ as a function that solves (10) but that has limited volume, $\varrho(t)\in L^1(\mathbb{R}^D\times\mathbb{R}^N)$ and $\varrho(t)\geq 0$ for $t>0$. In these conditions, if the support of the initial datum ϱ_0 is contained in $\mathcal{S}^D\times\mathbb{R}^N$, it can be shown as well that then also the support of $\varrho(t)$ is contained in $\mathcal{S}^D\times\mathbb{R}^N$. The following result then holds:

Theorem. *Assume that $\Phi_i, v\in C^{1,1}(\mathbb{R}^D\times\mathbb{R}^N)$, $\sigma\in C^{2,1}(\mathbb{R}^D\times\mathbb{R}^N)$ are bounded together with their derivatives, with $\Phi_i\geq 0$ and $\sigma\geq\varepsilon>0$. Take ϱ_0 a Lipschitz-continuous, bounded function whose support is compact and contained in the interior of $\mathcal{S}^D\times\mathbb{R}^N$ such that $\varrho_0\geq 0$ and $\iint \varrho_0 dxdy = 1$. Then there exists a unique viscosity solution to (10). Moreover $\varrho(t)\in L^1(\mathbb{R}^D\times\mathbb{R}^N)$ and $\varrho(t)\geq 0$ for all $t>0$.*

In the final part of Chapter 5 we provide numerical simulations concerning the two strategist game Hawks vs Doves, as presented in the above examples. We perform various simulations for the probability density obtained both by a Monte-Carlo method starting from the stochastic system (8), and by a finite difference scheme based on the Fokker-Plank equation (9). We show how equilibrium of the standard replicator-mutator dynamics can be disrupted by the effect of either random motion or mutations. In some particular cases, the environment itself allows for the survival of the least aggressive species, as we can see in Figure 2, where the probability density $\varrho(t)$ is shown. In the simulation we have "fair mutations", that is, mutations do not favor one strategy in particular. In the figure, with a red line it is represented the expected value of Hawks for $t\to\infty$ in absence of mutations, with a blue line the same

expected value for $t \to \infty$ in the presence of mutations. In that case

$$\lim_{t \to +\infty} \mathbb{E}[x_t] \approx \frac{5}{9},$$

then the expected value of the proportion of Hawks for $t \to +\infty$ is lower than the one without mutations, that is $2/3$; we see that including the physical space can favor the persistence of Doves.

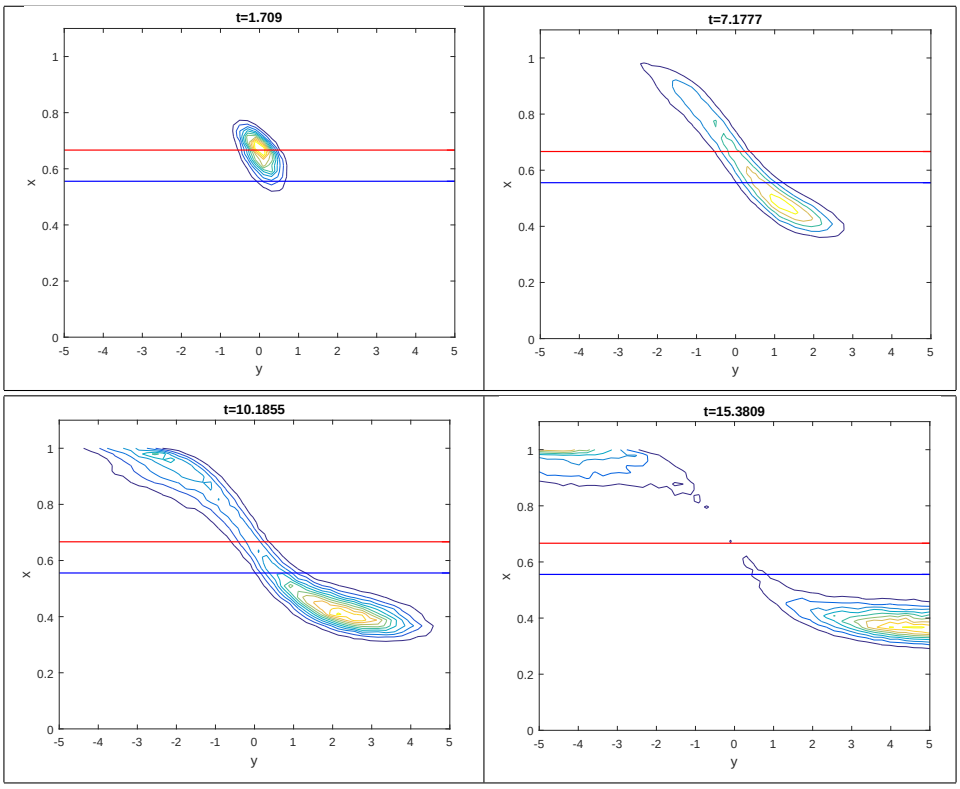


Figure 2: Point-type mutations plus deterministic and Brownian motion, The vector x evolves according to the game Hawks vs Doves with G function of y. The red line represents the initial value of Hawks and the expected value of Hawks for $t \to \infty$ in absence of mutations. The blue line is the same expected value for $t \to \infty$ in the presence of mutations. Brownian motion has $\sigma = 0.2$, the parameters of the jump process are $\lambda_{12} = \lambda_{21} = 0.2$, $\gamma_{12} = \gamma_{21} = 0.1$, that is one tenth of the population mutate each jump and we have "fair jumps".

The model presented and analyzed in the following Chapter 6 is based on model (6) with rare mutations, but adds the two main concepts of *difference between genotype and phenotype* and *potentially infinite genome space*. The GENOTYPE is the set of characteristics that define an individual, encoded in the DNA, the *information* of an organism, replicable and modifiable due to the random processes of mutations. A "type" of individuals in this model is a group that shares the same genotype. Each genotype expresses a very specific PHENOTYPE, and the same phenotype can be the expression of several different genotypes, on which natural selection acts. The main forces of Evolution, that are selection/adaptation and mutation, act then at different levels: the environment select the fittest individuals, acting on the phenotype, seen

in contest as *a function of the selective pressures*, while the source of evolutionary novelty, mutation, acts instead by randomly modifying the genotypes, producing new and never seen before ones.

Genotypes. In Nature, a genome can, in the first instance, be represented by a sequence of finite length of nitrogenous bases, or a finite string of elements in the alphabet $\{A, C, T, G\}$. Mutations can alter the individual bases, leaving the length intact, or modify for instance the code in a more substantial way, through deletions or insertions of bases. In this case the space of the genomes is represented by all the possible strings of finite length in the alphabet $\{A, C, T, G\}$, that has infinite elements, i.e.

$$\bigcup_{n\in\mathbb{N}} \{A, C, T, G\}^n .$$

In general, we define the POTENTIAL GENOME SPACE as a set $\mathcal{G}$, countable, whose elements g are the GENOMES. The unknown variable of the complete model, N, indicates a *population of genomes*, a vector (infinite) that contains in position $k \in \mathbb{N}$ the number n^k of individuals of genotype g_k. Only a finite set of genomes are present in the population, therefore only a finite number of elements is greater than zero (Figure 3). Formally:

Definition. In the space of all sequences of scalars $N = \left(n^1, \dots n^k, \dots\right)$, $n^k \in \mathbb{R}$ such that $n^k = 0$ except for finitely many n, let $\mathcal{S}^\infty_{\mathbb{R}}$ be the subset of the successions with compact support:

$$\mathcal{S}^\infty_{\mathbb{R}} = \left\{ N = \left(n^1, \dots, n^k, \dots\right) : \, n^k \geq 0 \, \forall k; \, \sum n^k < +\infty; \, D := \left| \left\{ k : \, n^k > 0 \right\} \right| < +\infty \right\}$$

A POPULATION OF GENOMES is an element $N \in \mathcal{S}^\infty_{\mathbb{R}}$.

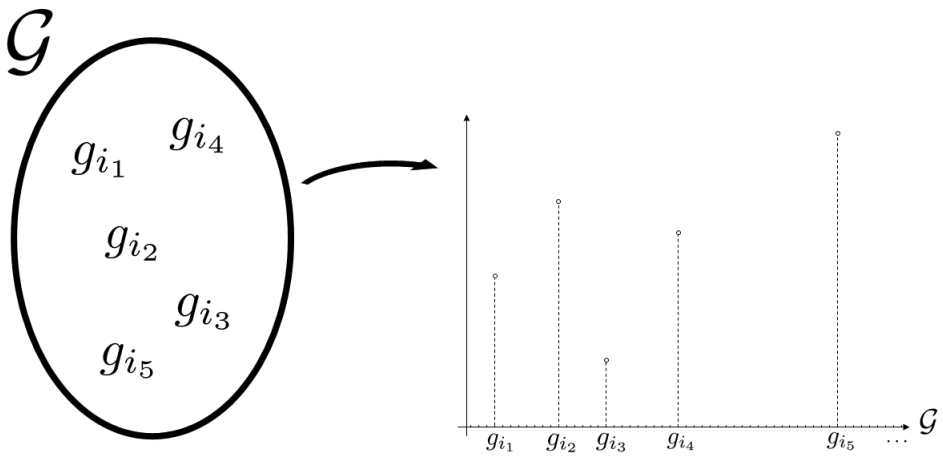


Figure 3: Starting from a genome space $\mathcal{G}$, a population of genomes N is defined: the genomes are shown, with their number. Only a finite number of genomes have number not null.

Phenotypes and Selection. Similar to the genotypes, a POTENTIAL PHENOTYPE SPACE is a set $\mathcal{F}$, whose element $f \in \mathcal{F}$ is a *phenotype*. Each genotype in $\mathcal{G}$ express a phenotype, therefore the two spaces $\mathcal{G}$ and $\mathcal{F}$ are linked. Such connection is given by a function, the GENOTYPE-PHENOTYPE MAP (GP) [36, 1]:

$$GP : \mathcal{G} \to \mathcal{F}.$$

A GP map is not injective, because different genotypes may generate the same phenotype. The set of genotypes that produce a given phenotype is called the *Neutral Space* of that phenotype.

The space of the phenotypes $\mathcal{F}$ is different from that of genotypes $\mathcal{G}$, and we want that selection acts only on the phenotypes. Selection can not distinguish two individuals that have the same phenotype, even if they have different genotypes; a phenotype is associated to a genotype through the GP-map, and a fitness function is to be defined associated to a phenotype. See Figure 4.

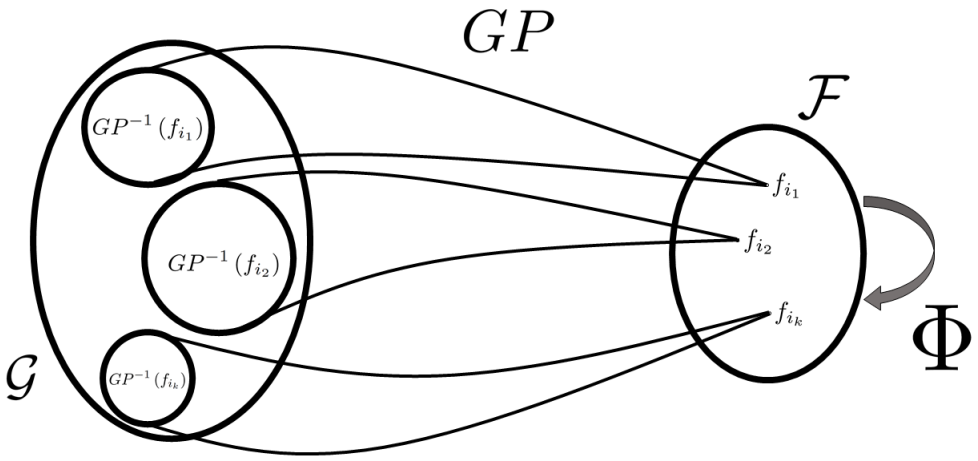


Figure 4: The phenotype-genotype map GP allows genotypes in $\mathcal{G}$ to express phenotypes in $\mathcal{F}$, and it is non-injective. Fitness function Φ acts exclusively on the phenotype space, ignoring $\mathcal{G}$.

Given the concepts of genotype, phenotype and fitness, it is possible to reformulate the equations of the replicator dynamics (1) for a population of genomes N.

First of all, model (1) studies densities of individuals (or assume the population constant in number). Since we want the mutation process leading to the formation of small groups of individuals, of very small absolute number (see next paragraph), we need to transform the Replicator dynamics into an equation for the number of individuals. This, following [44], can be done knowing that the Replicator dynamics for d types is equivalent to a Lotka-Volterra system with $d-1$ types. The next step is to add a term of logistic growth, then to rewrite the equations to take into account the space $\mathcal{S}^{\infty}_{\mathbb{R}}$. Overall, the differential system (1) is generalized,

for all $g \in \mathcal{G}$, with $N_0 \in \mathcal{S}_{\mathbb{R}}^{\infty}$, as:

$$N_t = N_0 + \int_0^t \tilde{a}\left(N_s\right) ds, \tag{11}$$

$$\tilde{a}\left(N\right) = \begin{cases} \frac{F(N)}{\sum_{j:\, g_j \in \mathcal{G}} n^j} n^k + n^k \left(\Phi_k(N) - \frac{1}{\sum_{j:\, g_j \in \mathcal{G}} n^j} \sum_{j:\, g_j \in \mathcal{G}} n^j \Phi_j(N) \right), \\ k \in \mathbb{N} \end{cases}$$

Mutation process. The model (11) *without mutations* is unnecessarily complex, since the presence of infinite space is irrelevant if an initial group of genomes N_0 can not explore $\mathcal{G}$. Mutation is the force that generates new genotypes, consequently increasing the dimensionality of the problem. Intuitively, we want the evolution of initial number of genotypes to change only according to selection and adaptation (the deterministic term given by $\tilde{a}$), until a mutation happens. A mutation will occur randomly in the population, with a temporal frequency proportional to the total number of individuals. When a mutational event occurs, a small group of individuals of a given genotype transforms into a new group, with a genotype similar to that of the progenitors. After the mutation, the process increase in dimensionality, and continue to evolve deterministically until the next mutation.

- *A mutation event occurs.* The process is based on a non-homogeneous Poisson of intensity proportional to the number of individuals in the population $\lambda = \lambda(N)$. The jumps happen therefore with frequency proportional to the number of individuals.

- *Identification of the mutants.* Mutations occur "blindly" by randomly selecting a small group of individuals with a fixed genome. The choice is proportional to the frequency of the different types in the population at the time of the mutation.

- *Choice of the new genome.* Once the genome that changes has been selected, it explores the space $\mathcal{G}$, changing in a new element of $\mathcal{G}$, which is "near" to him. If we imagine to follow the evolutionary history of a genome, we would see it moving between one vertex and another of a graph on $\mathcal{G} \times \mathcal{G}$, "Evolution is a trajectory through sequence space [73]"; so we define, fixed $g \in \mathcal{G}$, the probability $\Pi_{g.h}$ that, if a mutation involves g, it will mutate in $h \in \mathcal{G}$. The choice of the new genome must therefore take into account the graph on $\mathcal{G}$, shown in Figure 5

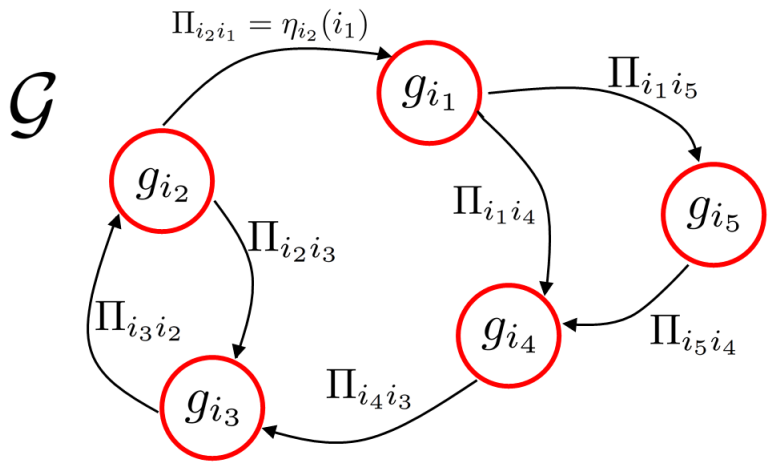


Figure 5: Example of graph on the genome space $\mathcal{G}$. The arrows have the weights Π_{jk} and represents mutation jumps.

$$N_t = N_0 + \sum_{n:T_n \leq t} 1_{\left[0,\lambda\left(N_{T_n^-}\right)\right]}(z_n) \sum_{j \in \mathbb{N}} 1_{I_j\left(N_{T_n^-}\right)}(u_n)\gamma_j\left(n^j_{T_n^-}, v^j_n\right) \tag{12}$$

Combining the deterministic term (11) and the stochastic one (14) we can formulate the complete process, whose existence and uniqueness is proved at the end of Chapter 6.

$$\begin{cases} N_t = N_0 + \int_0^t \tilde{a}\left(N_s\right) ds + \int_0^t \int_E K\left(N_{s^-}, \xi\right) \mathcal{N}\left(ds, d\xi\right), \quad t > 0 \\ N_0 \in \mathcal{S}^{\infty}_{\mathbb{R}}. \end{cases} \tag{13}$$

Using the construction (12) for the jump process, and a finite difference algorithm for the deterministic equation, it is then possible to implement the system numerically (13). An example, able to summarize the main characteristics of the model, is given by the *Fibonacci GP-map* ([36]):

By imposing these characteristics to the jump process, we obtain:

$$N_t = N_0 + \int_0^t \int_E K\left(N_{s^-}, z, u, v\right) \mathcal{N}\left(ds, d(z, u, v)\right);$$

where the mark space E is $\mathbb{R}^+ \times [0,1] \times (\mathcal{G} \times \mathcal{G})$ and K has the following form:

$$K\left(N, z, u, v\right) = 1_{[0,\lambda(N)]}(z) \sum_{k \in \mathbb{N}} \left\{ 1_{I_k(N)}(u) 1_{\{\alpha(k)\} \times \mathcal{G}}(v) \gamma_k\left(n^k, v\right) \right\}. \tag{14}$$

As stated in (4), if we let $(T_n)_{n \in \mathbb{N}}$ be a homogeneous Poisson process with intensity $\lambda_{max} := max(\lambda(N))$, and we define three collections of independent random variables, $(z_n)_{n \in \mathbb{N}} \sim Unif\left(0, \lambda_{max}\right)$, $(u_n)_{n \in \mathbb{N}} \sim Unif\left(0,1\right)$, and $\left(v^j_n\right)_{j,n \in \mathbb{N}} \sim Unif\left(\{\alpha(j)\} \times \mathcal{G}\right)$, then we can rewrite the jump process as

- a genome in $\mathcal{G}$ represents a simplified DNA string, in which the nitrogenous bases are reduced from four to two, 0 and 1. When a mutation occurs in the genome, the string can have a *punctual mutation*, i.e. an element chosen randomly within it changes from 0 to 1 or vice versa; the string can have an *insertion*, i.e. a 0 or a 1 is added into the genome, in a random position;
- A genome g of length L, $g \in \{0,1\}^L$, expresses a phenotype $f \in \mathcal{F}$, which is a binary string of length shorter than L, obtained in the following way:
  - starting with the first digit the sequence is considered "coding" until a "stop codon" is encountered,
  - after the stop codon the sequence is considered "non-coding".

  See Figure 6.

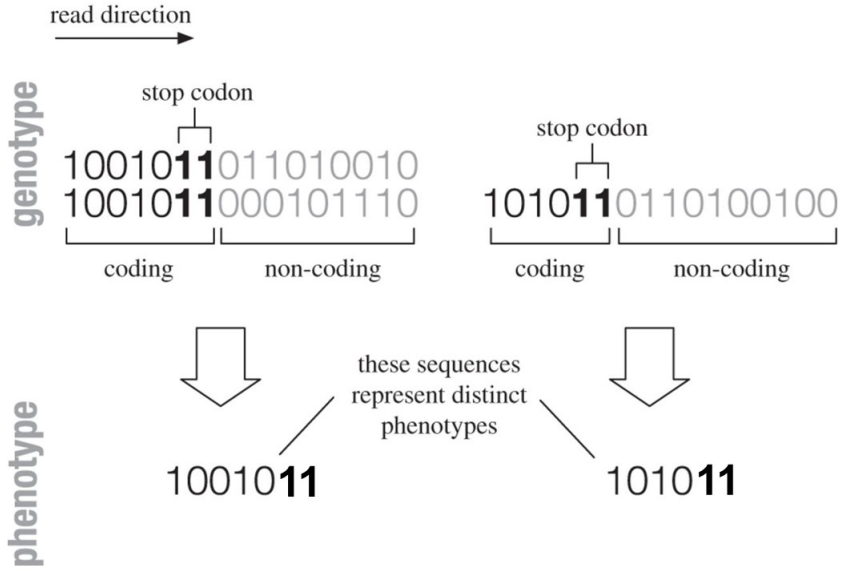


Figure 6: Three examples of genotype sequences, which map two different phenotypes in the *Fibonacci genotype-phenotype map*. Reading from the left the sequence is regarded as "coding" up to the first occurrence of the "stop codon" sequence 11. Thereafter the sequence is regarded as "non-coding". Each possible coding sequence represents a different phenotype, whereas the non-coding sequence leaves the phenotype entirely unaffected. Figure by [36].

It is then possible to define fitness functions on the phenotypes, and show how the presence of a *structure* in the genotype space can change the trend of the process.

In conclusion, in Chapters 5 and 6 we present the two main and innovative models of population dynamics with random mutations. In the first, published in the article [5], the presence of a heterogeneous environment, capable of influencing fitness functions, changes the relative proportions of individuals. In particular cases the physical space can favor the persistence of low fitness strategy. The second presents the evolution over time of a population of genomes described by a system of ordinary stochastic differential equations. The importance

of the difference in action between mutations, which modify the genotypes, and selection, which through fitness functions changes the proportions between phenotypes, is underlined. The presence of a potentially infinite space of available genomes makes sure that variants of individuals with characteristics never seen before can be generated.

To study the relationship between genotype and phenotype, in this thesis the choice was made to have *functions* as GP-map, so that each genotype can express a single phenotype; this assumption is a simplification of reality, as there exists situations in which a single genome can give rise to different phenotypes, e.g. due to different environments. A further development could then concern an expansion of the concept of GP-map. Another path of research could create a link between the two models, presented imagining the influence of a heterogeneous environment in the genotype and phenotype model; we could therefore obtain a model in which various populations, interacting with each other through gene exchanges, give rise to emigration and invasion of more fitted individuals.

In Chapter 9, we present the article *A new restart procedure for combinatorial optimization and its convergence* ([83]), written after a collaboration with G. Sebastiani, that deals with combinatorial optimization, probability and numerical simulation. Although not directly related to previous differential models, many algorithms are inspired by biological phenomena, such as the Ant Colony Optimization, or the Genetic Algorithms:

- Ant Colony Optimization [23] simulates the behavior of an anthill in optimizing the path that connects nest and food. Given a search domain, like the Hamiltonian cycles on a graph, a certain number of ants are released and start to explore it. During the exploration, each ant releases a trace of pheromone, which highlights the path it has made, and which is able to attract the other ants. Initially, a certain number of ants is placed. Then, in each construction step, each ant moves, based on a probabilistic decision, to a point of the search domain it has not yet visited. This probabilistic choice is biased by the pheromone trail. Ants prefer paths which are close and connected by arcs with a high pheromone trail. After all ants have completed the tour construction, the pheromone trails are updated, first by lowering them by a constant factor (evaporation) and then by allowing the ants to deposit pheromone on the path they have visited. If a path is not chosen, its associated pheromone trail decreases exponentially; this enables the algorithm to "forget" bad choices over time. The better the ant's tour is, the more pheromone is received. In general, paths which are used by many ants and are shorter will receive more pheromone and therefore will more likely be chosen in future iterations of the algorithm.

- Genetic Algorithms [65, 82] are inspired by the theory of Evolution, simulating replication, mutation, crossing-over and selection. Genetic algorithms have been used by many researchers as a tool for search and optimization. A finite collection of "genomes" (the candidate solutions) is given in a model "world", and a fitness function on this collection is defined, which has to be maximized. Usually, the number of genomes is very large prohibiting a complete search. Genetic algorithms provide a probabilistic way to conduct a search in the space of genomes, given a suitable encoding of the candidate solutions is strings of symbols. A genetic algorithm comprises three phases: *mutation*, *crossover* and *fitness selection*. These are applied cyclically and iteratively until some

condition is satisfied. The model most commonly investigated is the genetic algorithm with a binary alphabet, where genomes take the form of bit strings.

The article proposes a new iterative procedure to optimize the *restart time* of a meta-heuristic algorithm to solve *combinatorial optimization problems*. Solving a combinatorial optimization problem (COP) generally consists in finding an element, in a finite research domain, that minimizes a fitness function. The typical example of COP is the *Traveling Salesman Problem* (TSP), i.e. finding the shortest path that connects a given number of points, or cities. Even if a solution of a COP always exists, finding it often requires a very high computational cost. Two different search algorithms can be used: exact or heuristic.

Those of *heuristic* nature guarantee an exact solution only if infinite time is available, or a suboptimal solution in finite time. The article deals with meta-heuristic algorithms (MHA), which are heuristic algorithms independent of the particular COP considered. These algorithms therefore present a methodology for solving a problem, without specifying the problem in detail. Genetic algorithms and Ant Colony Optimization are of this type.

A natural problem concerns the convergence of this meta-heuristic algorithms. Given their stochastic nature, they can be studied by probability theory; unfortunately, even when convergence is guaranteed, algorithms are often too slow to be used in practice. It is in this contest that the restart procedure is used, which consists in repeating a very high number of times independent instances of a given MHA, randomly initialized. The best solution, among those produced, is chosen. With fixed available computational resources, the problem is to understand how many independent instances to produce, and how long to run them. In the article a new algorithmic procedure for the restart is proposed, and then it is applied to instances of (TSP), with hundreds and thousands of cities. As the basic algorithm, Ant Colont Optimization is used (in a version called Max-Min Ant System).

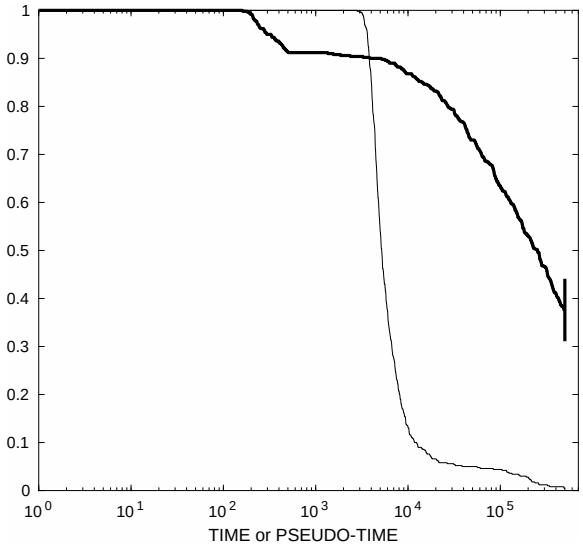


Figure 7: A TSP instance with 532 cities. The estimated failure probability as function of time for the standard Ant System (thick line) and the Restart Procedure (thin line). The time axis is in logarithmic scale. The failure probability of the Restart Procedure decreases by several orders of magnitude compared to the standard algorithm.

Using the new procedure we can show how the failure probability decreases by several orders of magnitude compared to the standard algorithms (see Figure 7), for equal computational cost. Therefore, given a certain computation resource, by applying this procedure, we are far more confident that the result obtained is a solution of the problem analyzed. Further developments of this procedure could regard preserving its performance and decreasing the computational cost, working with the control parameters along the iterations.

Part I

Short introduction to Game Theory, Evolutionary Dynamics, and Stochastic processes

Chapter 1

Evolutionary games

1.1 Population dynamics

Population dynamics is a branch of ecological mathematics that studies, through ordinary differential equations, the interactions between species in an ecosystem and the evolution (in mathematical sense) of their density over time. In natural ecosystems, thousands of different species interact by creating the complex structure of the biosphere, making it difficult, if not impossible, to provide an all-encompassing mathematical description. In the following we will analyze classical and simple models for the study of one or more populations that, as a first approximation, consider the effects of the struggle for the exploitation of resources (such as food) and of the competition between different species living in the same environment.

Malthusian growth model. The simplest model for a single species is the one proposed by Thomas Robert Malthus in the early 1800s, In his book *An Essay on the Principle of Population* [55] the author predicts a difficult future for humanity, assuming that there is an exponential growth of individuals, but linear growth of food resources. Let R the growth rate of a population, due to factors such as the abundance of food and the average number of offspring, and suppose that this rate is constant. We can then formulate an equation for the quantity $\dot{x}/x$, which represents the variation in time of the number of individuals of a population indicated with $x(t)$ and obtain the Malthusian growth equation

$$\dot{x} = Rx,$$

which has the solution $x(t) = \mathrm{e}^{Rt}x_0$, where x_0 is the amount of initial population. The behavior of such a solution is trivial, with an exponential growth for $R > 0$, a decrease until extinction for $R < 0$, a constant population for $R = 0$.

Despite the simplicity of the model, Malthus' work, mainly a treatise on sociology and economics, inspired Charles Darwin and Alfred Wallace in the development of the theory of natural selection. The Malthusian growth for $R > 0$ can be used as a first approximation for the analysis of the dynamics of a population that invades a new territory in the absence of constraints that limit its expansion, factors that, eventually, become increasingly important, limiting the speed of growth.

Logistic growth. The initial exponential growth of a population sooner or later slows down, as there are fewer resources for each individual and therefore a lower growth rate. In the simplest case we can consider this rate decreasing linearly as a function of x, that is $k(1-x)$. We are assuming in this case that the value 1 is the carrying capacity of the population, which can not grow further. This choice leads to the logistic equation, first named by P.F. Verhulst [92]:

$$\begin{cases} \dot{x} = kx(1-x), \\ x(0) = x_0 \geq 0. \end{cases}$$

The behavior of the solution is easy to analyze; the problem can be integrated, has $x = 0$ and $x = 1$ as equilibriums and the explicit solution is:

$$x(t) = \frac{x_0 \mathrm{e}^{kt}}{1 - x_0 (1 - \mathrm{e}^{kt})}.$$

Assuming a small initial density x_0 and a positive growth rate k, an initial sudden increase in density occurs, followed by a gradual slowdown, with $x(t)$ that reaches the carrying capacity.

Lotka-Volterra. Different models consider the interaction between several species in the same environment; among them the classical example is the system proposed indipendently by Vito Volterra and Alfred Lotka in 1920s [54, 93]. Volterra, in the years following the First World War, finds himself analyzing the dynamics of predator and prey among the fish caught in the Adriatic Sea. The amount of predators had increased considerably compared to previous years. The war had naturally influenced fishing in those years, but Volterra's question is how and why there had been a more favorable increase in predators rather than prey. The dynamic of the system proposed considers two populations, one of prey, one of predators and build a set of differential equations. The density of the prey is indicated by the function $x(t)$, while with $y(t)$ we indicate that of predators. It is assumed that the growth rate of the prey population, in the absence of predators, is given by a Malthusian growth, with parameter α, positive. The growth rate decreases linearly as a function of the density of predators y. Predators are based on the abundance of prey to live and, in the absence of them, the function y decreases exponentially by a factor $-\gamma$ (with γ positive). However, the growth rate of predators increases linearly with respect to the amount of prey x. The Lotka-Volterra system reads:

$$\begin{cases} \dot{x} = (\alpha - \beta y)\, x, \\ \dot{y} = (-\gamma + \delta x)\, y, \\ x(0) = x_0 \in \mathbb{R}^+, \\ y(0) = y_0 \in \mathbb{R}^+. \end{cases}$$

The system, that has $\mathbb{R}^+ \times \mathbb{R}^+$ as an invariant set, admits two distinct equilibria, one trivial with $(x_0, y_0) = (0, 0)$ and the other

$$(x_0, y_0) = \left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta} \right).$$

Because of the function $U(x, y) = \gamma \log x - \delta x + \alpha \log y - \beta y$ is constant on the solution of the equation, the dynamics is periodic; the density of prey and predators will oscillate, with

amplitude and frequency of the oscillations depending on the initial conditions. Thanks to this model, Volterra is able to give an answer to the problem of the Adriatic Sea: fishing reduces the rate of increase of the prey and it augments the rate of decrease of the predators. The average density of predators becomes smaller, the prey's one larger, than the unperturbed state: ceasing to fish leads to an increase of predators and a decrease of prey.

1.2 Game theory

Game theory was created, as a unique field, by John von Neumann in 1920s with the paper *On the Theory of Games*, followed by the book with Oskar Morgenstern in 1944 [71, 67]. They wanted a mathematical theory that was able to study human behavior when strategic and economic decisions were to be considered. Infact, Game Theory deals in general with mathematical techniques to analyze situations, called GAMES, in which two or more individuals, called PLAYERS, make decisions that influence their own and others' fitness. Players are always supposed to be rational decision-makers, i.e. able to make decisions that maximize their expected fitness. Game theory was further developed by John Nash in 1950, who in his PhD thesis [69], defined what is now called Nash equilibrium.

To illustrate the kind of problems that the theory tries to treat we can consider a classic and well-known example since 1950, the PRISONER'S DILEMMA.

Two criminals are accused of committing a crime. The investigators arrest them both and close them in two different cells, preventing them from communicating. Each prisoner is given the opportunity either to betray the other by testifying that the other committed the crime, or to cooperate with the other by remaining silent. The offer is:

- if one betrays, but the other remains silent, the betrayer avoid the penalty and is set free; the other will however be sent 5 years in prison;
- if both criminals each betray the other, they will have a penalty discount and will both be sent 4 years in prison;
- if they remain silent, both will be sent 2 years to prison, because they are incriminable exclusively for minor offenses.

Suppose the two promised to remain silent in case of arrest. They are now locked up in two different cells and wonder if the promise will be kept on the other. There is therefore a dilemma: to betray or not to betray. Game theory helps us solve these kinds of problems.

Definition 1. A normal form game is given by:

- a finite set G of *players* (in the above dilemma are the two prisoners);
- a finite set of pure strategies $S_1 \times \ldots \times S_G$ that players can implement. We are interested in $S_i = S_j$ for every i, j, i.e. when each player draws from the same set of strategies S, of cardinality $|S|$ (the two criminals can *betray* or *remain silent*);
- for each player in G a fitness function, or *payoff*

$$U_i : S_1 \times \ldots \times S_G \mapsto \mathbb{R},$$

(for each of his choices, the criminal risks more or less years in prison). Note that the function U_i does not depend only on the behavior of the player i but also from the strategy that the others implement.

A game between two players can be represented by a payoff matrix $\mathcal{U}$, with the number of rows equal to the number of possible strategies S, with $u_{i,j}$ payoff of a player who uses the i strategy against one of strategy j. In the case of two player strategies one has:

		s_1, strategy player 2 $\downarrow$	s_2, strategy player 2 $\downarrow$
s_1, strategy player 1	$\rightarrow$	$u_{1,1}$	$u_{1,2}$
s_2, strategy player 1	$\rightarrow$	$u_{2,1}$	$u_{2,2}$

or

$$\mathcal{U} = \begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix}.$$

Once a game has been modeled, we can try to find out which strategies represent plausible outcomes, assuming players are rational and pursuing their own interests. We can now give the concept of of Nash equilibrium.

Definition 2. A pure Nash equilibrium of a game with G players is a set of strategies $\bar{s}_1, \ldots, \bar{s}_G$ such that

$$U_i(\bar{s}_1, \ldots, \bar{s}_G) \geq U_i(\bar{s}_1, \ldots, s_i, \ldots, \bar{s}_G),$$

for each player $i = 1, \ldots, G$ and for each choice of strategy $s_i \in S_i$.

The interpretation is as follows: if a game accepts at least one Nash equilibrium, each player has at his disposal at least one strategy from which he has no interest in moving away if all the others do not change their strategy. In fact, if the player i plays any other strategy at his disposal other than $\bar{s}_i$, while everyone else continues to play with $\bar{s}_j$, it could only worsen its utility or, at most, leave it unchanged. So if players reach a Nash equilibrium, no one can improve their result by modifying only their own strategy, and is therefore bound to the choices of others. Since this applies to all players, however, if there is Nash equilibrium and it is unique, it represents the solution of the game, as none of the players have an interest in changing strategy.

Example. The prisoner's dilemma can be studied with a payoff matrix $\mathcal{U}$ for the strategies C strategies (cooperating) and D (defectors) in the form

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix}.$$

If they both cooperate they get a benefit of R (reward); if both betray, everyone receives P (punishment); if one of the two cooperates and the other betrays, the cooperator gets S (payoff of the stupid) while the betrayer gets T (temptation payoff). Parameter are sorted with $T > R > P > S$. The strategy (C, C) is the one that guarantees the greatest gain (the least total number of years in prison) but it is not a Nash equilibrium. If one of the two, convinced of the honesty of his companion decides to betray, avoid the penalty, and condemn him to

many years in prison. The game admits a unique Nash equilibrium, the strategy (D, D), where are both defectors, choice that causes more damage to both than the alternative (C, C) but which guarantees less risk and fewer years in prison for the individual. The dilemma is a very interesting game also at the biological level, so much so that in the articles *Prisoner's dilemma in an RNA virus* and *Escape from Prisoner's Dilemma in RNA Phage* Φ6 [90, 91], it is explained how more viruses that infect the same cell generate conflicts, the creation and the intracellular diffusion of viral products and their consequent sharing, allowing the creation of cooperation and selfishness behaviours: a viral genotype that synthesizes large quantities of products is actually cooperating, on the contrary a genotype which synthesizes it less, but which develops the ability to subtract most of the shared product, is a defector/selfish. In the articles a slightly different payoff matrix is proposed, equivalent to the previous one,

$$\begin{pmatrix} 1 & 1-s_1 \\ 1+s_2 & 1-c \end{pmatrix},$$

with all the constants positive. In this way

$$1+s_2 > 1 > \begin{cases} 1-s_1, \\ 1-c. \end{cases}$$

With c the cost of the interaction between two selfish people is indicated, with s_1 the cost paid by a cooperator for colliding with a selfish, with s_2 the selfish gain on that same occasion.

In a game for two agents (1 and 2) and two pure strategies (S_1 and S_2) the parameters to obtain a pure Nash equilibrium can be calculated. We consider the generic payoff matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

- The pair $(1, 1)$ is a pure Nash equilibrium if $a \geq c$;
- The pair $(2, 2)$ is a pure Nash equilibrium if $d \geq b$;
- The pair $(1, 2)$ is a pure Nash equilibrium if $b \geq d$ and $c \geq a$, the same for the couple $(2, 1)$, in fact, applying the definition,

However, a game does not always admit a single pure Nash equilibrium, and not always an equilibrium exists at all, as in the following example, the popular game Rock-Paper-Scissors.

Example. We take into account the three players game, with three pure strategies R (Rock), P (Paper) and S (Scissors), with payoff matrix $\mathcal{U}$

$$\begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

The game is perfectly symmetrical, with every strategy that is better against one and worse against the other: none of them is a pure Nash equilibrium.

Players have no way to "interpolate" between their actions: the selection is between a pure strategy s_i, or a different strategy s_j. This constraint can be relaxed by allowing players to select a probability distribution on the choices. For example, a player could select the strategy s_i with probability 1/3 and the action s_j with probability 2/3 (in general, it could select the strategy s_i with probability $p_i \geq 0$, provided that $\sum p_i = 1$). Such strategies are mixed, in contrast to the deterministic or pure strategies and are uniquely determined by the distribution $p = (p_i)_{i=1,\dots|S|}$, with p_i probabiliy to play the strategy s_i; they are therefore usually indicated with the vector. The concept of mixed Nash equilibrium can therefore be defined:

Definition 3. A set of mixed strategies $\bar{m}_1, \dots, \bar{m}_G$ is a Nash equilibrium if no player can unilaterally improve his expected utility by switching to a different mixed strategy, so it is a set of strategies $\bar{m}_1, \dots, \bar{m}_G$ such that, for each player $i = 1, \dots, G$ and for each choice of mixed strategy m_i of that player:

$$U_i(\bar{m}_1, \dots, \bar{m}_G) \geq U_i(\bar{m}_1, \dots, m_i, \dots, \bar{m}_G).$$

Since mixed strategies are a generalization of pure ones, it is not difficult to show that every pure Nash equilibrium is also a Nash equilibrium. The inverse is not true and in fact there are games without pure equilibrium but which admit mixed equilibria. Indeed, much more is true: every finite game admits at least one Nash equilibrium, and this is the fundamental result that John Nash presents in his PhD thesis [69].

Example. The previous game Rock-Paper-Scissors admits a mixed Nash equilibrium, one in which both players choose the strategy $\bar{p} = (1/3, 1/3, 1/3)$ that is, they randomly choose one of the three pure strategies. Consider for example player 1 and check that when player 2 uses strategy $\bar{p}$, the former has no incentive to play strategies other than $\bar{p}$. If the second player uses $\bar{p}$ and the first uses $q = (q_1, q_2, q_3)$, the expected payoff for player 1 becomes

$$q_1 \cdot \frac{1}{3} \cdot (0) + q_1 \cdot \frac{1}{3} \cdot (-1) + q_1 \cdot \frac{1}{3} \cdot (+1) +$$

$$q_2 \cdot \frac{1}{3} \cdot (+1) + q_2 \cdot \frac{1}{3} \cdot (0) + q_2 \cdot \frac{1}{3} \cdot (-1) +$$

$$q_3 \cdot \frac{1}{3} \cdot (-1) + q_3 \cdot \frac{1}{3} \cdot (+1) + q_3 \cdot \frac{1}{3} \cdot (0) = 0.$$

So the expected utility $u_1(q, \bar{p})$ is 0, regardless of the choice of q_1, q_2, q_3. The same is true when player 1 chooses q and player 2 chooses $\bar{p}$. Since when both players choose the mixed strategy $\bar{p}$, the utility is $u_1(\bar{p}, \bar{p}) = u_2(\bar{p}, \bar{p}) = 0$, then $\bar{p}$ is a Nash equilibrium:

$$U_1(\bar{p}, \bar{p}) \geq U_1(q, \bar{p}),$$

$$U_2(\bar{p}, \bar{p}) \geq U_2(\bar{p}, q).$$

1.3 Evolutionary games

Game theory is applied in various fields of knowledge such as Economics, Social Sciences and Philosophy and can also be successfully applied in the study of Evolution. An example of behavior that tends to maximize a payoff can indeed be found in Darwinian selection models:

in a universe where increasing disorder is a physical law, complex organisms (including social organizations) can survive only if they behave in a way that tends to increase their chances of survival and reproduction ([8]). Then an argument on evolutionary selection suggests that individuals tend to maximize the expected value of some measure of natural survival and reproductive fitness, otherwise they are replaced ([57]).

The theory of evolutionary games is born as an application of the theory of games to Biology. It was John Maynard Smith, professor of biology at the University of Sussex, who introduced the notion of evolutionarily stable strategy and successfully applied the theory of games to evolution. As he himself says in the introduction to his book *Evolution and the Theory of Games* [57], paradoxically, game theory is better applied to biology than to economic behavior, to study which it had been invented. There are two reasons that justify this assertion: the first is that the theory requires that there exists a utility function, which in human applications could turn out to be an artificial concept, whereas in biology one can associate it with Darwinian adaptability, the reproductive success of an individual or of a certain genotype. The second and perhaps most important reason is that in seeking the solution of a game the concept of human rationality is replaced by evolutionary selection. The advantage, says Maynard Smith, is that there are various reasons to expect the population to evolve towards stable states while there are many doubts about the rationality of human behavior.

Before going into the details of the meaning of evolutionarily stable equilibrium and its connection with the Nash equilibria, we must review the way in which we understand game theory. As mentioned, in traditional game theory we refer to the cognitive abilities of the players, we assume that the players are rational decision makers, who are able to understand that they are playing and are consciously trying to maximize their payoff, trying to predict the opponent's moves; these facts are used to justify why players should choose strategies that give rise to Nash equilibria. The theory of evolutionary games, on the other hand, starts from a very different concept, it assumes that players are individuals with inheritable strategies coded in their genome, and that they have no control over the strategies they use and generally do not even know they are players. Individuals reproduce and are subject to the forces of natural selection, so in general an individual will always implement the same strategy and it is the entire population that changes the type of strategy over time, moving towards more suitable strategies, clashing with many other individuals and repeating the same game over and over again.

History of ritual struggles, Hawks vs Doves

A phenomenon that shows the power of game theory in biological field is that expressed with the name Hawks and Doves (game introduced by Maynard Smith and Price in 1973 [60]), and starts from a consideration of zoological nature: despite almost all the complex animals (including us) have bilateral symmetry, as if the left side of the body had been created by mirroring the right, the males of the violinist crabs of the genus Uca have one of the two claws larger than the other. In everyday life the overdeveloped claw is a handicap for the crustacean, forced to live with such an encumbrance. About the social interactions between males, the violent clashes that would justify possession of such a powerful weapon are a very rare event, while almost all comparisons by territory and mating result in ritualized struggles in which the giant claw is shaken up and down rhythmically and they never pass to action

except in extreme cases; in those moments, however, usually the more peaceful of the two gives up almost immediately. Why did violinist crabs develop such ritual struggles if they have such big claws? The most aggressive animals could simply attack mercilessly at the expense of the peacemakers, win most of the clashes and transmit aggression to offspring. Why are not they all aggressive?

Suppose there are two behaviors (strategies) in the colony of Uca, aggressive and peaceful: during a fight the aggressives soon get tired of the ritual dances, and attack the opponents without mercy; face to face with another aggressive, they fight until exhaustion. The peacefuls, on the other hand, tend to avoid the struggles; face to face with another peaceful they begin a simulated ritual battle until one goes away; face to face with an aggressive always run away. Usually the term Hawks is used to refer to the aggressives, Doves for the peacefuls. Whatever the reason for the fight, the contenders are faced according to the following payoff matrix for the Hawks and Doves strategies [44]:

$$\mathcal{U} = \begin{pmatrix} \frac{G-C}{2} & G \\ 0 & \frac{G}{2} \end{pmatrix}.$$

Winning a fight increases the fitness by a quantity $G > 0$ (Gain); losing a physical fight gives a price to pay in terms of fitness equal to C (Cost). The cost C is chosen greater than the gain G; abandoning the fight does not produce variations in the fitness. Then,

- If two Doves meet, they perform ritual dances, then one retires. The winner gets G, the loser nothing, and on average their fitness increase by $G/2$;
- A Dove that meets a Hawk escapes by not getting anything, the Hawk takes everything for itself, earning G;
- A Hawk that meets a Hawk attacks it, fights until one of the two gives up. The winner's fitness increases by G, while the loser's loses C, with an average of $(G-C)/2$ (the price to pay for losing a physical battle is greater than the gain, so the average in this case is negative and results in a loss of fitness).

This game has no pure Nash equilibrium, because in a population composed almost entirely of Doves a few Hawks would dominate because, meeting only peaceful animals, have a gain almost equal to G. In a population of aggressive Hawks, on the other hand, it would be the Doves to win, because their fitness would remain the same while the Hawks would lose fitness fighting with each other.

Over time, evolution leads to a balance with a mixed population of $\frac{G}{C}$, value for which the fitness of the Hawks is the same as the Doves, solving the problem of ritual struggles: paradoxically, if the animals are very heavily armed and therefore a physical battle would lead to serious damage (C is very large), the G/C ratio it is very small and therefore the population at equilibrium consists of many Doves and a few Hawks. In the case of the Uca, the big claw is synonymous of great damage during a fight and then eventually a peaceful population has evolved, that use to simulate struggles, a social custom in which to show others the disproportionate limb to remember that a possible fight would not lead to anything good.

Evolutionary stable strategies

We reformulate the language of game theory in terms that we use in the evolutionary case [44]:

- An individual's strategy is a complete action plan for every situation that may arise; this completely determines the behavior of the player, that is the actions that the individual takes at any moment of the game, for every possible game history up to that point. A strategy consists of possible moves and probability distributions on those moves, which represent how often every move is implemented.
- A pure strategy defines a specific move or action that the player follows in every possible situation in the game. A pure strategy gives a complete and deterministic explanation of how a player plays his game. The set of a player's strategies is the set of strategies that are available to that player. In the game of Hawks and Doves an individual who has the pure Hawk behavior always attacks every individual he finds.
- A mixed strategy is a probability distribution on pure strategies, allowing the player to choose a pure strategy based on a given probability distribution. In this context, mixed strategies have a more natural interpretation: an individual who plays Hawks and Doves with strategy, for instance, $(1/2, 1/2)$ randomly chooses, with the same probability, what behaviour to implement in each of the fight he faces.

We assume that the fights take place between two individuals, that we have a game with N pure strategies $S_1, \ldots, S_N$ and that a player can use mixed strategies; these consist of playing $S_1, \ldots, S_N$ with probability $q_1, \ldots, q_N$, with q_i non-negative and with sum equal to one. A strategy is therefore a point q in the simplex

$$\mathcal{S}_N = \left\{ q = (q_1, \ldots, q_N) \in \mathbb{R}^N : \, q_i \geq 0; \sum_{i=1}^{N} q_i = 1 \right\}.$$

The elements of the canonical basis of $\mathbb{R}^N$, in the form $e_i = (0, \ldots, 1, \ldots, 0)$ represent the pure strategies, particular mixed strategies in which with probability 1 is chosen the strategyS_i. Let $u_{i,j}$ be the payoff for a player who uses pure strategy R_i (represented by $e_i \in \mathcal{S}_N$) against a player using pure strategy R_j (the point $e_j \in \mathcal{S}_N$). The matrix with N rows and N columns $\mathcal{U} = (u_{i,j})$ is the payoff matrix. If U is the utility function we therefore have that

$$U(R_i, R_j) = e_i \cdot \mathcal{U} e_j = u_{i,j}.$$

An individual with pure strategy R_i, against an individual with mixed strategy $q \in \mathcal{S}_N$ has utility function equal to

$$U(R_i, q) = e_i \cdot \mathcal{U} q = (\mathcal{U} q)_i = \sum_j u_{i,j} q_j.$$

The payoff of a player with mixed strategy p against one of mixed strategy q is

$$U(p, q) = p \cdot \mathcal{U} q = \sum_{i,j} u_{i,j} p_i q_j.$$

The concept of Nash equilibrium in this context has a different interpretation,

Definition 4. A strategy $\bar{p} \in \mathcal{S}_N$ is called (symmetric) Nash equilibrium if it is the best answer to itself, that is if for each strategy $p \in \mathcal{S}_N$, $p \neq \bar{p}$,

$$p \cdot \mathcal{U}\bar{p} \leq \bar{p} \cdot \mathcal{U}\bar{p}.$$

If equality is not valid, the equilibrium is strict.

Note that in this formulation we do not have a couple of strategies as in the traditional definition, but a single element of $\mathcal{S}_N$, we are imagining that the whole population adopts the equilibrium strategy. Consider a large population of players: if everyone adopted the strategy $\bar{p}$, a strict Nash equilibrium, any new strategy that occurs occasionally (for example through mutation processes) is penalized and therefore can not supplant $\bar{p}$. We can not however assume that every equilibrium is strict, even in the game Hawks and Doves is not, so we can not even assume that every Nash equilibrium is resistant to the invasion of a dissident minority. The following definition is therefore necessary, introduced and coined by Maynard Smith in 1970s [58],

Definition 5. The strategy $\bar{p} \in \mathcal{S}_N$ is evolutionarily stable (ESS) if for each strategy $q \in \mathcal{S}_N$ different from $\bar{p}$, and for each $\varepsilon > 0$ smaller than an appropriate threshold $\bar{\varepsilon}(q) < 1$, this inequality holds:

$$q \cdot \mathcal{U}\left(\varepsilon q + (1-\varepsilon)\bar{p}\right) < \bar{p} \cdot \mathcal{U}\left(\varepsilon q + (1-\varepsilon)\bar{p}\right). \tag{1.1}$$

The term $\varepsilon q + (1-\varepsilon)\bar{p}$ indicates a population in which a dissident minority q appeared, therefore a strategy is an ESS if, when every member of the population has adopted it, no dissident behavior can invade it, being resistant to the occasional appearance of other strategies

The equation (1.1) can be reformulated as

$$(1-\varepsilon)\left(\bar{p} \cdot \mathcal{U}\bar{p} - q \cdot \mathcal{U}\bar{p}\right) + \varepsilon\left(\bar{p} \cdot \mathcal{U}q - q \cdot \mathcal{U}q\right) > 0,$$

then $\bar{p}$ is an ESS if and only if the following two conditions are met:

- *equilibrium condition*
$$q \cdot \mathcal{U}\bar{p} \leq \bar{p} \cdot \mathcal{U}\bar{p} \, for \; all \, q \in \mathcal{S}_N;$$
- *stability condition*
$$if \; q \neq \bar{p} \; and \; q \cdot \mathcal{U}\bar{p} = \bar{p} \cdot \mathcal{U}\bar{p}, \; then \; q \cdot \mathcal{U}q < \bar{p} \cdot \mathcal{U}q.$$

The condition of equilibrium is the definition of Nash equilibrium; an ESS is something more than just the property of being the best answer to itself, which does not guarantee the security of not being invaded, because it allows the existence of another strategy that is a better answer (as in Hawks and Doves, with a population of Doves that can be invaded by Hawks, or vice versa). The evolutionarily stable strategies are characterized by the following lemma (proved in [44]),

Lemma 6. *The strategy $\bar{p} \in \mathcal{S}_N$ is an ESS if and only if, for each $q \neq \bar{p}$ near $\bar{p}$ in $\mathcal{S}_N$ it holds:*

$$\bar{p} \cdot \mathcal{U}q > q \cdot \mathcal{U}q.$$

1.4 Replicator Dynamics

Replicator Dynamics, introduced by Taylor and Jonker [88], describes the evolution of the frequencies of the strategies in a population. Assume that the population is divided into n types $E_1, \ldots, E_n$ with frequencies $x_1, \ldots, x_n$. The fitness f_i of an individual of type E_i is a function of the composition of the population, i.e. the state $x = (x_1, \ldots, x_n)$ on the simplex $\mathcal{S}_n$ (different from the space of strategy $\mathcal{S}_N$, presented above). If the population is very large, and if the generations pass continuously from one to the other, we can assume that the state $x(t)$ evolves on $\mathcal{S}_n$ as a differentiable function of time t. The rate of increase of the population fraction of type E_i is indicated by $\dot{x}/x$, ratio between variation over time $\dot{x}$, and quantity x. According to Darwinism, if the rate of increase is a measure of evolutionary success, we can express this success as the difference between fitness $f_i(x)$ of E_i and the average fitness of the population,

$$\bar{f}(x) = \sum x_i f_i(x).$$

These choices lead to the equation called *replicator equation* ([44]),

$$\begin{cases} \dot{x}_i = x_i \left(f_i(x) - \bar{f}(x) \right), \\ i = 1, \ldots, n. \end{cases} \tag{1.2}$$

We observe that the simplex $\mathcal{S}_n$ is invariant for the dynamics of (1.2): if $x \in \mathcal{S}_n$, then $x(t) \in \mathcal{S}_n$ for all time $t \in \mathbb{R}$. In fact the sum $S = x_1 + \ldots + x_n$ resolves:

$$\dot{S} = (1 - S)\, \bar{f},$$

that has $S(t) = 1$ as a solution (so if the solution of (1.2) remains on the space $\sum x_i = 1$). If $x_i(0) = 0$ then $x_i(t) = 0$ for every t, then the faces of the simplex $\mathcal{S}_n$ are barriers that make $\mathcal{S}_n$ invariant.

Interesting is the case when the fitness f_i is linear, in this case there is a matrix $A = (a_{i,j})$ such that $f_i(x) = (Ax)_i$, then $\bar{f}(x) = x \cdot Ax$; the equation replicator turns into

$$\begin{cases} \dot{x}_i = x_i \left((Ax)_i - x \cdot Ax \right), \\ i = 1, \ldots, n. \end{cases} \tag{1.3}$$

The equilibrium points in int$\mathcal{S}_n$ are the solutions of

$$(Ax)_1 = \cdots = (Ax)_n, \quad \sum x_i = 1,$$

for $x_i > 0$. In general, except in degenerate cases, there is almost one solution. This case is closely related to the evolutionary games theory developed earlier because it is the equation that is naturally obtained with populations playing a game with a payoff matrix $\mathcal{U}$. So define a game in normal form with N pure strategies $R_1, \ldots, R_N$ and a payoff matrix $\mathcal{U}$. A (mixed) strategy is a point in $\mathcal{S}_N$, then the n types $E_1, \ldots, E_n$ of the population correspond to n points $p^1, \ldots, p^n \in \mathcal{S}_N$. The state of the population is defined by the frequencies x_i of types E_i and the dynamics happens on $\mathcal{S}_n$, with fitness matrix $a_{i,j} = p^i \cdot \mathcal{U} p^j$ and then an equation (1.3) with

$$f_i(x) = \sum_j a_{i,j} x_j = (Ax)_i.$$

The values of N and n are different, but it is convenient to create a parallel between the pure strategies $R_1, \ldots, R_N \in \mathcal{S}_N$ with payoff array $\mathcal{U}$ and the types of the population $E_1, \ldots, E_n \in \mathcal{S}_n$ with the fitness matrix A . In particular, we say that a point $\bar{x} \in \mathcal{S}_n$ is a (symmetrical) Nash equilibrium if

$$x \cdot A\bar{x} \leq \bar{x} \cdot A\bar{x},$$

for each $x \in \mathcal{S}_n$ and it is an evolutionarily stable state (ESS) if

$$\bar{x} \cdot Ax > x \cdot Ax, \tag{1.4}$$

for all $x \neq \bar{x}$ near $\bar{x}$ in $\mathcal{S}_n$.

The following results on the replicator equation hold (details in [44]),

Theorem 7. *Let $\bar{x} \in \mathcal{S}_n$ and consider the equation (1.3).*

- *If $\bar{x}$ is a Nash equilibrium of the game with matrix A, then $\bar{x}$ is an equilibrium point of the equation.*
- *If $\bar{x}$ is the ω-limit of an orbit $x(t)$ in $\mathcal{S}^n$, then $\bar{x}$ is a Nash equilibrium.*

Theorem 8. *If $\bar{x} \in \mathcal{S}^n$ is an ESS for the game with matrix A, then is an asymptotically stable rest point for (1.3).*

Proof. This result is obtained choosing a function P such that

$$P(x) = \prod x_i^{\bar{x}_i}.$$

This function has a single maximum in $\mathcal{S}^n$ at the point $\bar{x}$. Infact:

$$\log P(x) - \log P(\bar{x}) = \log \prod x_i^{\bar{x}_i} - \log \prod \bar{x}_i^{\bar{x}_i} = \sum \bar{x}_i \log x_i - \sum \bar{x}_i \log \bar{x}_i,$$

setting as usual "$0 \log 0 = \infty \log 0 = 0$" we can write

$$= \sum \bar{x}_i \log \frac{x_i}{\bar{x}_i} = \sum_{\bar{x}_i > 0} \bar{x}_i \log \frac{x_i}{\bar{x}_i};$$

we can now use Jensen's inequality: *if φ is a convex function defined on some interval I, then*

$$\varphi\left(\sum p_i y_i\right) \leq \sum p_i \varphi(y_i) \tag{1.5}$$

for all $y_1, \ldots, y_n \in I$ and $p_i > 0$ $\forall i$, $\sum p_i = 1$, with equality if and only if $y_1 = y_2 = \ldots = y_n$. We apply this with $y_i = x_i/\bar{x}_i$, $\varphi = -\log$, $p_i = \bar{x}_i$ and we obtain:

$$\sum_{\bar{x}_i > 0} \bar{x}_i \log \frac{x_i}{\bar{x}_i} \leq \log \sum_{\bar{x}_i > 0} \bar{x}_i \frac{x_i}{\bar{x}_i} \leq \log \sum x_i = \log 1 = 0.$$

Then

$$\log P(x) - \log P(\bar{x}) \leq 0$$

$$P(x) \leq P(\bar{x})$$

with equality if and only if $x = \bar{x}$.

If $P > 0$ (true for all $x \in \mathcal{S}^n$ with $x_i > 0$ when $\bar{x}_i > 0$) then we have

$$\frac{\dot{P}}{P} = D(\log P) = D(\sum \bar{x}_i \log x_i) = \sum_{\bar{x}_i > 0} \bar{x}_i \frac{\dot{x}_i}{x_i} = \sum \bar{x}_i \left((Ax)_i - x \cdot Ax\right) = \bar{x} \cdot Ax - x \cdot Ax.$$

Since $\bar{x}$ is evolutionarily stable (1.4) implies $\dot{P} > 0$, for each $x \neq \bar{x}$ near $\bar{x}$. Then the function P is a strict local Lyapunov function for the linear replicator dynamics, and all the orbits that start near $\bar{x}$ converge to $\bar{x}$. □

In what follows, we do not explicitly specify the different types in the population, assuming they correspond to the pure strategies of the game itself, and that therefore $n = N$ and $\mathcal{U} = A$, i.e. each game is associated with a "pure strategy dynamics".

Two-strategy games

We conclude this chapter by solving the generic problem of the the equation with two-strategy game, and with some examples. From the equation (1.3), for $n = 2$, holds:

$$\begin{cases} \dot{x}_0 = \left[f_0(x_0, x_1) - \bar{f}(x_0, x_1)\right] x_0, \\ \dot{x}_1 = \left[f_1(x_0, x_1) - \bar{f}(x_0, x_1)\right] x_1, \\ x_0 + x_1 = 1. \end{cases}$$

The last condition allows to further simplify the problem, considering only the second equation and substituting

$$\begin{bmatrix} x \to x_1 \\ 1 - x \to x_0 \\ s = (f_0 - f_1) \end{bmatrix}$$

$$\dot{x} = \left[f_1 - \bar{f}\right] x = \left[f_1 - f_1 x - f_0 (1 - x)\right] x = (f_1 - f_0)\, x(1 - x),$$

$$\dot{x} = -sx(1 - x). \tag{1.6}$$

with $s = f_0 - f_1$. The functions f_0 and f_1 can be chosen according to a payoff matrix.

Constant fitness. If f_0 and f_1 are constant, suppose $f_0 > f_1$, then s is constant and positive. Since individuals of type 0 always have a higher fitness than the one of type 1, we expect an evolution that leads the population to become all of type 0.

$$\begin{cases} \dot{x} = -sx(1 - x), \\ x(0) = x_0. \end{cases}$$

The solution is:

$$x(t) = \frac{x_0 \mathrm{e}^{-st}}{1 - x_0 \left(1 - \mathrm{e}^{-st}\right)},$$

e it has as $x = 0$ and $x = 1$ has equilibria, one is attractive, one is repulsive, function of the sign of s.

Hawks and Doves. If with x we intend the fraction of population of Hawks, the equation is

$$\dot{x} = \frac{C}{2}\left(\frac{G}{C} - x\right) x\,(1-x)\,.$$

We are interested in knowing qualitatively the behavior. The righthand side is null for $x = 0$ and $x = 1$, that is, the population does not vary when there are no Hawks or no Doves, but ther is also the equilibrium, for $x = \frac{G}{C}$ (between 0 and 1 because the cost of the fight C is greater than the gain G), which corresponds to the value for which the fitness of the Hawks is equal to that of the Doves, different from the other two because attractive: the trajectory will be pushed towards $\frac{G}{C}$, for every initial x different from 0 and 1. Whatever the initial population is, even with an infinitesimal presence of Doves (or Hawks), after a sufficiently long time it will stabilize around the value $\frac{G}{C}$.

RNA Virus. Referring to [90, 91], we consider the case of the RNA virus phage $\Phi 6$; a population is divided into cooperators (C, in the variable x), viruses that synthesizes large quantities of product, and defectors (D, in the variable y), viruses that develops the ability to subtract most of the shared product, with a payoff matrix $\mathcal{U}$ equal to

$$\begin{pmatrix} 1 & 1-s_1 \\ 1+s_2 & 1-c \end{pmatrix},$$

with all constant positive. The equation becomes

$$\begin{cases} \dot{x} = \left[f_c(x,y) - \bar{f}(x,y)\right] x, \\ \dot{y} = \left[f_d(x,y) - \bar{f}(x,y)\right] y, \\ x + y = 1, \end{cases}$$

with

$$f_c = \left[A \begin{pmatrix} x \\ y \end{pmatrix}\right]_1 = x + (1-s_1)y,$$
$$f_d = \left[A \begin{pmatrix} x \\ y \end{pmatrix}\right]_2 = (1+s_2)x + (1-c)y,$$
$$\bar{f} = x \cdot [x + (1-s_1)y] + y \cdot [(1+s_2)x + (1-c)y]\,.$$

By reducing the variables, calling $y = 1-x$, the equation for x becomes $\dot{x} = (f_c - f_d)\,x(1-x)$, that is

$$\dot{x} = [x(s_1 - s_2 - c) - (s_1 - c)]\,x(1-x).$$

In addition to the equilibria in 0 and 1 there could also be a mixed equilibrium:

$$\bar{x} = \frac{s_1 - c}{s_1 - c - s_2},$$

- if $s_1 > c$ then $\bar{x} < 0$ or $\bar{x} > 1$ (based on the denominator sign). This is the case in which the price to pay for a confrontation with a defector is greater than that against a cooperator. The function s is negative in $[0,1]$ and pushes the solutions towards 0, a population of only defectors. In this case the game is a *prisoner's dilemma*;

- if $s_1 < c$, then $\bar{x} \in [0, 1]$. When you pay less to fight against a cooperator you have an interesting case. The function s is positive in $(0, \bar{x})$ and negative in $(\bar{x}, 1)$, therefore solutions are pushed towards equilibrium with both strategies, a mixed polymorphism. This is no more a prisoner's dilemma, but a *Hawk-Dove game.*

In the articles is showed that, starting from a situation where defectors are spreading in the population, as expected from the Prisoner's Dilemma [90], the game can change and evolve in a mixed polymorphism one. If the cooperators are allowed to grow under absence of competitive interactions, these evolved phagi should coexist in a mixed polymorphism with evolved defectors. "Thus, phage $\Phi 6$ is unlikely to be permanently trapped in a Prisoner's Dilemma. Rather, evolution of Prisoner's Dilemma is likely to be a local phenomenon for populations experiencing high rates of coinfection. If evolved cooperators were to enter these populations, the locally adapted defectors would face a mixed polymorphism at best [...]. In general, the transition from a Prisoner's Dilemma to a Hawks and Dove game can be achieved either by selection for more cheating and the associated costs (i.e., increasing c) or by selection for decreased sensitivity to cheaters (i.e., decreasing s_1). The experiments indicate that the latter is a possibility in $\Phi 6$. Thus, the mixed polymorphism was achieved because levels of cooperation are variable in this system.*[91]*"

Chapter 2

Stochastic differential equations

All the models presented in this work are of a stochastic nature, involving Brownian motions and jump processes. For this reason we give in this chapter recalls of theory of probability and stochastic processes. See [26] for a more detailed treatment. We recall that a stochastic process on measurable space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection $X = \{X(t)\}_{t\geq 0}$ or $\{X_t\}_{t\geq 0}$ of random variables. For each point $\omega \in \Omega$ the function $t \mapsto X(t,\omega)$ is a trajectory of the process.

Let X_t be a stochastic process. Then

$$\mathcal{U}(t) = \mathcal{U}(X_s | 0 \leq s \leq t),$$

the σ-algebra generated by the random variable X_s for $0 \leq s \leq t$, is called *history* of the process until time $t \geq 0$. Then, let assume that $\mathbb{E}(|X_t|) < \infty$ for all $t \geq 0$. If

$$X_s = \mathbb{E}(X(t)|\mathcal{U}(s)) \text{ a.s. for all } t \geq s \geq 0,$$

then X_s is a *martingale*.

Every stochastic process $X(t)$ is a function not only of the time t, but also of $\omega \in \Omega$, $X(t,\omega)$. For simplicity of notation we write $X(t)$ or X_t, leaving the dependency from ω implicit when possible.

2.1 Wiener and Jump processes

Wiener process

A Wiener process is a stochastic process that describes Brownian motion, the random motion of particles suspended in a fluid, resulting from their collision with the fast-moving molecules in the fluid. This motion is named after the botanist Robert Brown, who studied it in 1827, while looking through a microscope at pollen immersed in water.

Definition. A real valued stochastic process $\{W_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is a *standard Brownian motion* or *Wiener process* if:

- $W_0 = 0$ almost surely;
- for each $t \geq s \geq 0$, the increments $W_t - W_s$ are distributed as a Gaussian random variable with expected value 0 and variance $t - s$;

- it is a process with independent increments, i.e. for each finite sequence of times$0 < t_1 < t_2 < \ldots < t_n$, the random variables $W_{t_1}, W_{t_2} - W_{t_1}, \ldots , W_{t_n} - W_{t_{n-1}}$ are independent;
- with probability one, the function $t \mapsto W_t$ is continue.

In particular, $\mathbb{E}(W_t) = 0$ and $\mathbb{E}\left(W_t^2\right) = t$ for all time $t \geq 0$.

There is result that more explicitly highlights the physical characteristics of a Wiener process:

Lemma. *Let $\{B(t)\}_{t\geq 0}$ be a real valued stochastic process, with $B(0) = x \in \mathbb{R}$,*

• with independent increments (the events occurring in successive time intervals have no memory of those that occurred previously);

• stationary, i.e. $B(t+h) - B(t)$ has the same law as $B(s+h) - B(s)$ for each pair of times (t, s) positive, for $h > 0$ (the movement depends only on the length of the time interval, not on the extremes);

• with continuous trajectories almost surely.

Then there are values b and σ such that B is a Brownian motion *of drift b, diffusion coefficient σ^2 and intial point x, i.e:*

$$B(t) = x + bt + \sigma W(t).$$

A stochastic process on $\mathbb{R}^n$, $W_t = (W_t^1, \ldots, W_t^n)$ is a n-dimensional standard Wiener process if W_t^k is a Wiener process on $\mathbb{R}$ and if the $\sigma-$algebras $\mathcal{F}(W_t^k | \, t \geq 0)$ are independent, for each $k = 1, \ldots, n$.

Construction of the stochastic integral

We can write a diffusion process X in the form

$$X_t = X_0 + \int_0^t b(s)ds + \int_0^t \sigma(s)dW.$$

as soon as we have a definition for

$$\int_0^T GdW,$$

for an appropriate class of stochastic processes G. The definition is not obvious since $t \mapsto W(t, \omega)$ is with infinite variation for almost every omega. In fact, we show that the stochastic integral can have different definitions based on the choice made in the approximations of its construction, with different choices that lead to different theories. The reference texts for this section are [26] for the diffusion processes, [49, 75] for jump processes.

Definition. Let W be a one-dimensional Wiener process. The Ito integral $\int_0^T WdW$ is defined as

$$\int_0^T WdW = \frac{{W_T}^2}{2} - \frac{T}{2}.$$

The stochastic integral is defined for a wide range of stochastic processes, which are adapted with respect to the history $\mathcal{F}_t$, i.e. $\mathcal{F}_t$-measurable for all time $t \geq 0$. The idea is that for every

time t, the random variable G_t depends only on the information available in the σ -algebra $\mathcal{F}_t$. A process G is *progressively measurable* if it is adapted to filtration $\mathcal{F}_t$ and it is jointly measurable with respect to the variables t and ω together. A progressively measurable process at any time "depends only on the past history of Brownian motion".

Definition. We call $\mathbb{L}^2(0,T)$ the class of progressively measurable real processes such that

$$\mathbb{E}\left(\int_0^T G^2 dt\right) < \infty,$$

then we define $\mathbb{L}^1(0,T)$ the class of process with

$$\mathbb{E}\left(\int_0^T |G|\, dt\right) < \infty.$$

The set on which to build Ito integral is $\mathbb{L}^2$. Using the above definitions is infact possible to define stochastic integrals for processes in $\mathbb{L}^2$, and the following holds:

Lemma. *For $G \in \mathbb{L}^2(0,T)$ Ito stochastic integral $\int_0^T G_s dW_s$ is well defined, characterized by the following properties, for almost every $\omega \in \Omega$,*

- $\mathbb{E}\left(\int_0^T G(s,\omega) dW_s\right) = 0$;
- *The integral function $I(t) = \int_0^t G dW$ defined for $0 \le t \le T$ is a martingale with respect to $\mathcal{F}_t$.*

Definition. Suppose that X is a stochastic process in the integral form

$$X_t = X_0 + \int_0^t b(s) ds + \int_0^t \sigma(s) dW.$$

with $b \in \mathbb{L}^1(0,T)$, $\sigma \in \mathbb{L}^2(0,T)$. This process is a *diffusion process*, and for $0 \le t \le T$ has *stochastic differential*

$$dX = b dt + \sigma dW.$$

Poisson processes and Markov Chains

In the following, to model mutations, we will use the concept of jump process, i.e. processes that have discrete movements, *jumps*, rather than continuous movements, such as the Wiener process. We recall here the definitions of counting process, Poisson process and present a simple numerical implementation of it.

Definition. Let $(T_n)_{n\in\mathbb{N}}$ be a succession of positive random variables with real values, such that

$$\begin{cases} T_0 = 0, \\ T_n < T_{n+1}. \quad \text{if } T_n < +\infty \end{cases}$$

Then define the process $\{N_t\}_{t\ge 0}$:

$$N_t = \begin{cases} n, & \text{if } t \in [T_n, T_{n+1}) \\ +\infty. & \text{if } t \ge T_\infty \end{cases}$$

(N_t) is a *counting process* (for a fixed t, N_t counts the number of events in the interval $[0,t]$).

Examples of counting processes include Poisson processes:

Definition. A process $(N_t)_{t\geq 0}$ is a homogeneous Poisson process with intensity $\lambda > 0$ if it is a counting process, $N_0 = 0$, it has independent increments and

$$N_t - N_s \sim \mathcal{P}\left(\lambda(t-s)\right),$$

for all $0 \leq s \leq t$ and with $\mathcal{P}(\mu)$ a Poisson distribution with intensity μ, i.e.

$$\mathbb{P}\left(N_t - N_s = k\right) = \frac{(\lambda(t-s))^k}{k!}\mathrm{e}^{-\lambda(t-s)}, \quad k = 0, 1, \ldots$$

A Poisson process of intensity λ can be easily constructed considering $\{U_i\}_{i\in\mathbb{N}}$, sequence of independent random variables identically distributed as $Exp(\lambda)$, that is, with density function

$$f(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t}, & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The process takes the first jump at time U_1, the second after a time U_2 from the first and so on. In fact if we define the arrival times as $T_n = \sum_{i=1}^{n} U_i$ the counting process N associated is Poisson. If U is a random variable distributed as an exponential $Exp(\lambda)$, then $\mathbb{E}(U) = 1/\lambda$, so we have an expected number of $T \cdot \lambda$ jumps in a time interval $(0, T)$. Simulating a Poisson process is therefore simple, since it reduces to generate exponential variables starting from uniforms in $(0, 1)$. Let therefore T and U be random variables, $T \sim Exp(\lambda)$ and $U \sim Unif(0, 1)$, we calculate an appropriate function

$$g : [0, 1] \mapsto \mathbb{R},$$

such that $g(U) = T$. For $t \geq 0$

$$\mathbb{P}\left(g(U) \leq t\right) = \mathbb{P}\left(T \leq t\right) = 1 - \mathrm{e}^{-\lambda t},$$

$$\mathbb{P}\left(g(U) \leq t\right) = \mathbb{P}\left(U \leq g^{-1}(t)\right) = \mathbb{P}\left(U \in [0, g^{-1}(t)]\right) = g^{-1}(t),$$

so we choose g such that $g^{-1}(t) = 1 - \mathrm{e}^{-\lambda t}$ i.e.

$$T = g(U) = -\frac{\log\left(1 - U\right)}{\lambda}.$$

Because of U is uniform on $[0, 1]$ $(1 - U) \sim Uniform(0, 1)$, so we can consider $\bar{g}$,

$$T = \bar{g}(U) = -\frac{\log U}{\lambda}.$$

Markov Chains. Poisson processes are some of the simplest examples of continuous time Markov chains. Since during the presentation of the model of Chapter 6 we will use an idea taken from the construction of a continuous time Markov chain with discrete states, we recall the most important definitions we will need [72].

Definition. Let I be a countable set. Each $i \in I$ is called a *state* and I is called *state-space*.

- the quantity $\lambda = (\lambda_i : i \in I)$ is a distribution on I if $0 \leq \lambda_i \leq 1$ for all $i \in I$ and $\sum_i \lambda_i = 1$;
- a matrix $P = (p_{ij})$, $i, j \in I$ is stochastic if every row $\{p_{ij} : j \in I\}$ is a distribution;
- A stochastic process X_n, $n \in \mathbb{N}$ is a Markov chain with initial distribution λ and transition matrix P, Markov(λ, P) if
  1. X_0 has distribution λ, i.e. $\mathbb{P}(X_0 = i_0) = \lambda_{i0}$;
  2. for $n \geq 0$, conditional on $X_n = i$, X_{n+1} has distribution $\{p_{ij} : j \in I\}$ and is independent of $X_0, \ldots, X_{n-1}$, i.e.
$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = p_{i_n i_{n+1}}$$

For which it concerns continuous-time chains:

Definition. Let I be a state space.

- A Q-matrix on I is a matrix $Q = (q_{ij})$, $i, j \in I$ satisfying the following conditions:
  1. $0 \leq -q_{ii} < \infty$ for all i; we will write q_i as an alternate notation for $-q_{ii}$;
  2. $q_{ij} \geq 0$ for all $i \neq j$;
  3. $\sum_j q_{ij} = 0$ for all i;

Each off-diagonal entry q_{ij} gives the value we interpret as the rate of going from i to j. The numbers q_i are the rate of leaving i. The basic data for a continuos-time Markov chain on I are given in the form of a Q matrix. From a matrix Q is possible to obtain the *jump matrix* $\Pi = (\pi_{ij})$, $i, j \in I$, defined as:

$$\pi_{ij} = \begin{cases} q_{ij}/q_i & \text{if } j \neq i \text{ and } q_i \neq 0 \\ 0 & \text{if } j \neq i \text{ and } q_i = 0, \end{cases}$$

$$\pi_{ii} = \begin{cases} 0 & \text{if } q_i \neq 0 \\ 1 & \text{if } q_i = 0, \end{cases}$$

Here is the definitions of a continuous-time Markov and jump chain.

Definition. Define first the jump chain Y:

- Define a discrete-time Markov(λ, Π) Y_n and variables $S_1, S_2, \ldots$ that describe holding times in each of the states of Y_n, as independent exponential random variables of parameters $q(Y_0), \ldots, q(Y_{n-1})$ respectively.
- A right-continuous process X_t, $t \geq 0$ on I is a *Markov chain with initial distribution* λ *and generator matrix* Q if it has Y_n as jump chain and $S_1, \ldots, S_n$ as holding times.

We can construct such a process as follows: let Y_n, $n \geq 0$ be discrete-time Markov(λ, Π) and let $T_1, T_2, \ldots$ be independent exponential random variables of parameter 1, independent of Y_n. Set $S_n = T_n / q(Y_{n-1})$, $J_n = S_1 + \ldots + S_n$ and

$$X_t = \begin{cases} Y_n & \text{if } J_n \leq t < J_{n+1} \text{ for some } n \\ \infty & \text{otherwise.} \end{cases}$$

Then (X_t), $t \geq 0$ has the required properties.

Jump processes and Ito integration for jump processes

Let us consider now processes with jump terms and define the concept of stochastic integrals for this kind of processes too. Let$(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$ be a probability space with a filtration $\mathcal{F}_t$. A one-dimensional diffusion process *with jumps* has the form

$$X_t = X_0 + \int_0^t b(s)ds + \int_0^t \sigma(s)dW_s + J(t), \tag{2.1}$$

with

- $b \in \mathbb{L}^1(0,t)$ and $\sigma \in \mathbb{L}^2(0,t)$ for all t (a diffusion process);
- J a pure jump process, adapted to $\mathcal{F}_t$, right continuous and with $J_0 = 0$, with a finite number of jumps in each time interval $[0,T]$, constant between two successive jumps.

A clear expression can be found for pure jump processes, through the definition of Poisson random measure, that allows to define a new type of stochastic differential. See [49].

Definition. Let (E,ν) be a measurable space. A Poisson random measure $\mathcal{N}(dt,d\xi)$ on $\mathbb{R}^+ \times E$ is such that:

- for each interval $[x_1,x_2]$ and each measurable set $C \subset E$, $\mathcal{N}([x_1,x_2]\times C)$ is a Poisson random measure with parameter $(|x_2-x_1|\cdot\nu(C))$;
- if $[x_1,x_2]\times C$ and $[x_1',x_2']\times C'$ are disjoint, then $\mathcal{N}([x_1,x_2]\times C)$ and $\mathcal{N}([x_1',x_2']\times C')$ are independent.

In the case where E is of finite measure, $\nu(E) = \lambda < +\infty$, a formulation for a Poisson random measure can be obtained constructively. Infact, let us consider $\{\xi_i\}_{n\in\mathbb{N}}$, succession of random variables with values in E, independent and identically distributed with uniform law on (E,ν):

$$\mathbb{P}\left(\xi_i \in d\xi\right) = \frac{\nu(d\xi)}{\nu(E)}.$$

Then consider a Poisson process N_t of intensity $\lambda = \nu(E)$. With these ingredients $\mathcal{N}([x_1,x_2]\times C)$ is the measure that counts the number of points of the type $(T_i(\omega),\xi_i(\omega))$ within $[x_1,x_2]\times C$, that is

$$\mathcal{N}(dt,d\xi) = \sum_n \delta_{T_n}(dt)\cdot\delta_{\xi_n}(d\xi). \tag{2.2}$$

A pure jump process J, according to the definition of $\mathcal{N}(dt,d\xi)$ can be formulated in the following form:

$$J(t) = \int_0^t \int_E K(s,\xi)\mathcal{N}(ds,d\xi).$$

In $\mathcal{N}(dt,d\xi)$ are the temporal characteristics of the jumps, in $K(t,\xi)$ their amplitude. It is also useful to reformulate this form in the following way:

$$J(t) = \int_0^t \int_E K(s,\xi)\left[\mathcal{N}(ds,d\xi) - \nu(E)\frac{\nu(d\xi)}{\nu(E)}ds\right] + \int_0^t \int_E K(s,\xi)\cdot\nu(E)\frac{\nu(d\xi)}{\nu(E)}ds, \tag{2.3}$$

such that the second addend is in ds and the first verify the lemma below:

Lemma 9. *If, for all* $t > 0$, $\mathbb{E}\left[\int_0^t \int_E |K(s,\xi,\cdot)|\nu(d\xi)ds\right] < \infty$, *then*

$$\int_0^t \int_E K(s^-,\xi)\left(\mathcal{N}(ds,d\xi) - \nu(d\xi)ds\right) \tag{2.4}$$

is a martingale.

Definition. A process in the following form,

$$X_t = X_0 + \int_0^t [\dots]\, ds + \int_0^t [\dots]\, dW_s + \int_0^t \int_E [\dots]\left(\mathcal{N}(ds,d\xi) - \nu(d\xi)ds\right).$$

is a *semimartingale.*

The complete process (2.1), in martingale form, is

$$X_t = X_0 + \int_0^t \tilde{b}(s)ds + \int_0^t \sigma(s)dW_s + \int\int_E K(s,\xi)\left[\mathcal{N}(ds,d\xi) - \nu(E)\frac{\nu(d\xi)}{\nu(E)}ds\right] \tag{2.5}$$

where $\tilde{b}$ is the sum of the function b and the term in ds in (2.3).

2.2 Stochastic diffusive equations

Let W_t be a $m-$dimensional Wiener process and X_0 a $n-$dimensional random variable indepedente of W_t. Let then

$$\mathcal{F}(t) = \mathcal{F}\left(X_0, W_s, 0 \le s \le t\right)$$

be the $\sigma-$algebra generated from X_0 and from the history of the Wiener process up to time t.

For $T > 0$, let us define

$$b : \mathbb{R}^n \times [0,T] \longmapsto \mathbb{R}^n,$$

with $b = (b_1, \dots, b_n)$ and

$$\sigma : \mathbb{R}^n \times [0,T] \longmapsto \mathbb{M}^{n\times m},$$

with components

$$\sigma = \begin{pmatrix} b_{1,1} & \cdots & b_{1,m} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,m} \end{pmatrix},$$

deterministic assigned functions.

Definition. The stochastic process X_t, with values in $\mathbb{R}^n$ is the *solution of the stochastic differential equation*

$$\begin{cases} dX = b(X,t)dt + \sigma(X,t)dW, \\ X(0) = X_0, \end{cases}$$

for $0 \le t \le T$ if X_t is progressively measurable in respect of $\mathcal{F}(t)$, if $b(X,t) \in \mathbb{L}_n^1(0,T)$ and $\sigma(X,t) \in \mathbb{L}_{n\times m}^2(0,T)$ and if

$$X_t = X_0 + \int_0^t b(X_s,s)ds + \int_0^t \sigma(X_s,s)dW_s,$$

almost surely for $0 \le t \le T$. The coefficient b is the *drift,* σ is the *diffusion coefficient.*

Theorem 10. *Let us suppose that functions b, σ have not only the regularity required in the previous definition, but that the following conditions hold:*

- Lipschitz condition
$$|b(x_1,t)-b(x_2,t)|^2+|\sigma(x_1,t)-\sigma(x_2,t)|^2\leq L\cdot|x_1-x_2|^2,$$
for all times $0\leq t\leq T$, for a constant L and for all $x_1,x_2\in\mathbb{R}^n$;
- Growth condition
$$|b(x,t)|^2+|\sigma(x,t)|^2\leq L\cdot\left(1+|x|^2\right),$$
for all times $0\leq t\leq T$, for a constant L and for all $x\in\mathbb{R}^n$;
- Initial condition
$$\mathbb{E}\left(|X_0|^2\right)<\infty,$$
independent of $W(t)$.

Then a unique solution $X\in\mathbb{L}_n^2(0,T)$ of the differential equation exists:
$$\begin{cases}dX=b(X,t)dt+\sigma(X,t)dW,\\ X(t=0)=X_0.\end{cases}$$

Remark. A unique solution is unique almost everywhere, that is, if X and $\tilde{X}$ are both solutions of the same stochastic differential equation then
$$\mathbb{P}\left(X_t=\tilde{X}_t \text{ per ogni } 0\leq t\leq T\right)=1.$$

The proof of the theorem can be found on [26]. In conclusion we state a theorem on the dependence of the equation from its parameters ([26]).

Theorem 11. *Suppose that b^k, σ^k and X_0^k satisfy the hypotheses of the existence and uniqueness theorem all with the same constant L. Suppose they are solutions to problems*
$$\begin{cases}dX^k=b^k(X^k,t)dt+\sigma^k(X^k,t)dW,\\ X^k(t=0)=X_0^k.\end{cases}$$
Then let us assume that
$$\lim_{k\to\infty}\mathbb{E}\left(\left|X_0^k-X_0\right|^2\right)=0,$$
and that, for all $M>0$,
$$\lim_{k\to\infty}\max_{\substack{0\leq t\leq T\\ |x|\leq M}}\left(\left|b^k(x,t)-b(x,t)\right|+\left|\sigma^k(x,t)-\sigma(x,t)\right|\right)=0.$$
Then
$$\lim_{k\to\infty}\mathbb{E}\left(\max_{0\leq t\leq T}\left|X^k(t)-X(t)\right|^2\right)=0,$$
where X is the only solution of
$$\begin{cases}dX=b(X,t)dt+\sigma(X,t)dW,\\ X(t=0)=X_0.\end{cases}$$

Remark. In particular for almost every $\omega \in \Omega$ the random trajectories of the diffusive stochastic equation

$$\begin{cases} dX^\varepsilon = b(X^\varepsilon)dt + \varepsilon dW, \\ X^\varepsilon(t=0) = x_0, \end{cases}$$

converge uniformly on $[0,T]$, for $\varepsilon \to 0$ to the deterministic trajectory of

$$\begin{cases} \dot{x} = b(x), \\ x(0) = x_0. \end{cases}$$

Then, in the case of small random noises the stochastic trajectories are only a slight disturbance of the deterministic ones.

2.3 S.D.E.s with jump terms

Regarding the existence and uniqueness of solutions to the stochastic differential equation

$$\begin{cases} dX = b(X,t)dt + \sigma(X,t)dW + \int_E K(X,\xi)\mathcal{N}(dt,d\xi) \\ X(t=0) = X_0. \end{cases} \tag{2.6}$$

we can refer to Atreya [7] where an approach that takes advantage of the piecewise constant behaviour of the process is used.

First, we assume that b and σ satisfy conditions to guarantee the existence of the process $\tilde{X}$, satisfying (2.6) with $K = 0$, i.e. a pure diffusion process. We assume that, for all $x \in \mathbb{R}^n$, the process $\tilde{X}$ does not explode in finite time. Next we assume that

$$\lambda(x) = \nu\{\xi \in E, |K(x,\xi)| \neq 0\} < \infty$$

for all $x \in \mathbb{R}^n$ and λ is bounded on compacts.

These two hypoteses guarantee that the first jump time

$$\tau_1 := \inf\left\{t > 0, \int_0^t \int_E K\left(\tilde{X}_{s^-}, \xi\right)\mathcal{N}(ds,d\xi) \neq 0\right\}$$

fulfills $\mathbb{P}(\tau_1 > 0) = 1$ for all initial position $X_0 = x \in \mathbb{R}^n$. Thus, starting at $X_0 = x$, X_t evolves as the diffusion $\tilde{X}_t$ (with $\tilde{X}_0 = x$) for $t < \tau_1$ and then jumps to

$$X_{\tau_1} = \tilde{X}_{\tau_1} + K\left(\tilde{X}_{\tau_1}, \xi_1\right)$$

where (τ_1, ξ_1) is a point belonging to the Poisson random measure $\mathcal{N}(\cdot,\cdot)$ with the property that $\mathcal{N}((0,\tau_1) \times E) = 0$.

Next, starting at X_{τ_1}, the process evolves as the diffusion $\tilde{X}_t$ (with $\tilde{X}_0 = X_{\tau_1}$) up to a random time τ_2, then jumps and so on.

It is clear that under the above assumptions one can define the jump times $\tau_1, \tau_2, \ldots$ appropriately and the process X_t is well defined for all $t < \tau_\infty := \lim_{n\to\infty} \tau_n$. Thus this construction yields a solution to 2.6 up to explosion time τ_∞.

Assuming an uniform bound for $\lambda(x)$, i.e. $\lambda := \sup_{x\in\mathbb{R}^n} \lambda(x) < \infty$ is a sufficient condition to ensure that $\tau_\infty = \infty$ with probability one for all initial conditions $x(0) = x$

2.4 Numerical implementation of stochastic equations

For both types of processes considered, diffusion and jump, it is possible to construct numerically the stochastic integrals and consequently to implement methods to simulate stochastic differential equations. For what concerns the diffusion processes we will use results of article [43].

Brownian motion. We repeat for convenience the definitions of Wiener process, W_t, continuous, is a Wiener process if:

1. $W(t=0)=0$ with probability 1;
2. For $0 \leq s < t \leq T$ the increment $W_t - W_s$ is distributed as a Gaussian of expected value 0 and variance $t-s$, that is
$$W_t - W_s \sim \sqrt{t-s}N(0,1),$$
with $N(0,1)$ a standard Gaussian;
3. For $0 \leq s < t < u < v \leq T$ the increments $W_t - W_s$ and $W_v - W_u$ are independent.

For computational purposes it is convenient to consider a discrete Brownian motion, where W_t is defined only for a discrete set of values of t. Chosen the time interval $\delta t = T/N$ for an appropriate $N \in \mathbb{N}$, denote with W_j the value of W_{t_j} with $t_j = j\delta t$. The first condition states that $W_0 = 0$ and the subsequent ones tell us that

$$W_j = W_{j-1} + dW_j,$$

with $j = 1, \ldots, N$ and with dW_j a random variable independent of the previous, in the form $\sqrt{\delta t}N(0,1)$. Numerically, it is possible to generate only random variables uniforms in $[0,1]$, we can obtain Gaussian variables using the algorithm Box-Muller [11]:

Lemma. *Let Z_1 and Z_2 be two independent random variables, identically distributed, uniforms on $[0,1]$. Then the transformation*

$$V_1 = \sqrt{-2\log Z_1}\cos(2\pi Z_2),$$
$$V_2 = \sqrt{-2\log Z_1}\sin(2\pi Z_2),$$

transform the couple (Z_1, Z_2) in the pair (V_1, V_2), such that V_1 and V_2 are distributed as standard Gaussian variables, $N(0,1)$, independent of each other.

Stochastic differential dW and Euler-Maruyama. If we can simulate the process W, then the integral $\int dW$ can also be built, starting from the definition of Riemann sum and considering the approximation with rectangles

$$\int_0^T h(t)dW_t \simeq \sum_{j=0}^{N-1} h(t_j)\left(W_{t_{j+1}} - W_{t_j}\right).$$

The Euler-Maruyama method is a numerical method for simulating autonomous stochastic equations of the type

$$X_t = X_0 + \int_0^t f(X_s)ds + \int_0^t g(X_s)dW_s,$$

for f and g scalar functions with initial datum X_0, random variable. To simulate the equation, it is convenient to switch to compact notation

$$dX_t = f(X_t)dt + g(X_t)dW_t;\ X(t=0) = X_0. \tag{2.7}$$

First, the interval $[0,T]$ on which the problem is defined is discretized, so define $\Delta t = T/L$ for some L positive, and define $\tau_j = j\Delta t$ and $X_j = X_{\tau_j}$. It is convenient to choose the step Δt for the numerical method so that it is an integer multiple of the increment $\delta t = T/N$ of the discrete Brownian motion (defined on the N times $t_j = j\delta t$), i.e $\Delta t = R\delta t$, so that the set of points on which the Wiener process is calculated is a subset of the steps for the solution of the numerical stochastic equation. Calculating the equation in the points τ_j we get:

$$X_{\tau_j} = X_{\tau_{j-1}} + \int_{\tau_{j-1}}^{\tau_j} f(X_s)ds + \int_{\tau_{j-1}}^{\tau_j} g(X_s)dW_s,$$

and by approximating each term we obtain at the Euler-Maruyama method

$$X_j = X_{j-1} + f(X_{j-1})\Delta t + g(X_{j-1})\left(W_{\tau_j} - W_{\tau_{j-1}}\right),$$

for $j = 1, \ldots, L$. Brownian increments can be reformulated into

$$W_{\tau_j} - W_{\tau_{j-1}} = W(jR\Delta t) - W((j-1)R\Delta t) = \sum_{k=jR-R+1}^{jR} dW_k.$$

Note that in case the function g is null, the method is reduced to explicit Euler for ordinary equations.

Eulero-Maruyama with jumps. To take account of the jumps it is necessary to add terms to the Euler-Maruyama structure; focusing only on the new terms we have a S.D.E. of the form

$$dX_t = K(X_t)dN_t,$$

with N_t Poisson process of intensity λ. Using the definition of stochastic integral in $\mathcal{N}$,

$$N_t = \int_0^t K(X_s)\mathcal{N}(ds) = \sum_n \int_0^t K(X_s)\delta_{T_n}(ds),$$

then, calculating in the points discretized by Δt as in the previous case, we get the increase

$$\sum_n \int_{\tau_j}^{\tau_{j+1}} K(X_s)\delta_{T_n}(ds) = \sum_{\tau_j \le T_n \le \tau_{j+1}} K(X_{T_n}).$$

Assuming the interval Δt small we can approximate X_{T_n} with X_j and then write

$$\sum_{\tau_j \leq T_n \leq \tau_{j+1}} K(X_{T_n}) \simeq K(X_j) \cdot \#\left\{T_n : \tau_j \leq T_n < \tau_{j+1}\right\},$$

i.e. $K(X_j)$ multiplied by the number of jumps inside $[\tau_j, \tau_{j+1})$.

Recalling that $X(\tau_n)$ and X_n are random variables, it is necessary to define a concept of convergence order of a numerical method.

Definition. A numerical method for stochastic differential equations has a strong convergence order equal to γ if there is a constant C such that

$$\mathbb{E}\left(|X_n - X_\tau|\right) \leq C\Delta t^\gamma,$$

for all $\tau = n\Delta t \in [0, T]$ and for Δt small.

Euler-Maruyama method, with the functions f and g sufficiently regular, has strong convergence order $\gamma = 1/2$. It can be noted that there is a difference with the deterministic case, as choosing $g = 0$ and initial value X_0 constant, the expected value can be eliminated from the above inequality, which remains true also for $\gamma = 1$. One could try to improve the order of convergence by implementing more complex algorithms, such as the Milstein method or Runge-Kutta ([43]) methods.

Chapter 3

Stochastic processes and P.D.E.s

A stochastic process can be described in a deterministic way by means of the two Kolmogorov integro-partial differential equations: the backward one, related to *expected values*, and the forward one, related to the *probability density*. In this chapter we show how, starting from models of stochastic equations, it is possible to obtain these equations. We first establish a connection between infinitesimal generators of stochastic processes, i.e. partial differential operators that encode informations about the process, and parabolic equations describing macroscopic quantities of processes. We use Ito formula, a stochastic equivalent of Taylor's formula, to derive the Feynman-Kac equation (Kolmogorov backward), then we create a connection between the latter and the Fokker-Planck equation (Kolmogorov forward). The results presented below can be found in [75, 76, 49].

3.1 Ito formula

Diffusion processes

Ito formula is a fundamental formula of stochastic calculus that gives us the possibility to define the concept of infinitesimal generator, at the basis of Kolmogorov's theorems [26].

Theorem 12. *Suppose X a diffusive stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which solves the stochastic differential equation*

$$X_t = X_0 + \int_0^t b\,(X_s)\,ds + \int_0^t \sigma\,(X_s)\,dW_s,$$

$$dX = bdt + \sigma dW,$$

for $b \in \mathbb{L}^1(0,T)$, $\sigma \in \mathbb{L}^2(0,T)$. Let $u : \mathbb{R} \times [0,T] \mapsto \mathbb{R}$ be a continuous function with u_t, u_x and u_{xx} that exist and are continuous. Consider the process

$$Y_t = u(X_t, t).$$

Then Y solves the stochastic differential equation

$$Y_t - Y_0 = \int_0^t \left(u_t(X,s) + u_x(X,s)b + \frac{1}{2}u_{xx}(X,s)\sigma^2 \right) ds + \int_0^t u_x(X,s)\sigma dW_s. \tag{3.1}$$

$$dY = u_t dt + u_x dX + \frac{1}{2} u_{xx} \sigma^2 dt \tag{3.2}$$
$$= \left(u_t + u_x b + \frac{1}{2} u_{xx} \sigma^2 \right) dt + u_x \sigma dW,$$

Note that, since $X_t = X_0 + \int b ds + \int \sigma dW$, X has continuous trajectories almost surely and therefore for almost every $\omega \in \Omega$ the functions of times $u_t(X_t, t)$, $u_x(X_t, t)$, $u_{xx}(X_t, t)$ are continuous, then the integrals in ((3.1)) are well defined. The theorem can be generalized to the case of multiple stochastic processes with differential $dX^i = b^i dt + \sigma^i dW$, function $u : \mathbb{R}^m \times [0, T] \mapsto \mathbb{R}$ continuous with u_t,u_{x_i},u_{x_i,x_j} continuous for all $i, j = 1, \ldots, m$ to obtain a generalized Ito formula:

$$du(X^1, \ldots, X^m, t) = u_t dt + \sum_{i=1}^{m} u_{x_i} dX^i + \frac{1}{2} \sum_{i,j=1}^{m} u_{x_i,x_j} \sigma^i \sigma^j dt.$$

Obviously we can further extend the argument done in dimension one and get the Ito formula for processes in dimension $n \geq 1$, i.e. $dX = bdt + \sigma dW$ with $b \in \mathbb{L}^1_n(0, T)$, $\sigma \in \mathbb{L}^2_{n \times m}(0, T)$; for $i = 1, \ldots, n$

$$dX_i = b_i dt + \sum_{j=1}^{m} \sigma_{ij} dW_j.$$

Lemma 13. *Let u be a continuous function with domain in $\mathbb{R}^n \times [0, T]$, with u_t, u_{x_i}, $u_{x_i x_j}$ continuous for all i and j, then*

$$d\left(u(X_t, t)\right) = u_t dt + \sum_{i=1}^{n} u_{x_i} dX_i + \frac{1}{2} \sum_{i,j=1}^{n} u_{x_i x_j} \sum_{l=1}^{m} \sigma_{ij} \sigma_{jl} dt, \tag{3.3}$$

with derivatives computed in (X_t, t).

Ito for semimartingale

Also in the case of jumping processes, a Ito formula can be calculated:

Lemma 14. *[49] Let X be a semimartingale as in (2.3) and (2.5),*

$$dX = bdt + \sigma dW + dJ,$$

for $b \in \mathbb{L}^1$, $\sigma \in \mathbb{L}^2$. Define $u \in C^{2,1}\left(\mathbb{R} \times [0, +\infty)\right)$. Then the process $\{Y_t = u(X_t, t)\}_{t \geq 0}$ is a semimartingale and the following formula holds,

$$dY_s = u_t(X_s, s)ds + \mathcal{A}u(X_s, s)ds + dM^u_s, \tag{3.4}$$

with

$$\mathcal{A}u(x, t) = u_x(x, t)b(t, \omega) + \frac{1}{2} u_{xx}(x, t)\sigma^2(t, \omega) + \int_E \left[u(x + K(t, \xi), t) - u(x, t)\right] \nu(d\xi),$$

and M^u_t a martingale.

Proof. Let us consider the equation (2.5) between a jump and another, for $T_i \leq t < T_{i+1}$. We have

$$X_t = X_{T_i} + \int_{T_i}^{t} \tilde{b}(s)ds + \int_{T_i}^{t} \sigma(s)dW_s,$$

with the jump term that disappears because of the absence of jumps in the considered interval. In $[T_i, T_{i+1})$ Ito formula for diffusion processes can be used,

$$Y_t = Y_{T_i} + \int_{T_i}^{t} \left[u_s(X_s, s) + u_x(X_s, s)\tilde{b}(s) + \frac{1}{2}u_{xx}(X_s, s)\sigma^2(s) \right] ds$$

$$+ \int_{T_i}^{t} u_x(X_s, s)\sigma(s)dW_s. \tag{3.5}$$

This equation holds while $t = T_{i+1}^-$. Between times T_{i+1}^- and T_{i+1} there is a jump and X changes in value,

$$X(T_{i+1}) = X(T_{i+1}^-) + K(T_{i+1}, \xi_{i+1}, \omega).$$

Then, for Y,

$$Y(T_{i+1}) - Y(T_{i+1}^-) = \left[u\left(X(T_{i+1}^-) + K(T_{i+1}, \xi_{i+1}, \omega), T_{i+1}^- \right) - u\left(X(T_{i+1}^-), T_{i+1}^- \right) \right]. \tag{3.6}$$

We can then rewrite the increment for Y on the whole time interval $[0, T]$,

$$\begin{array}{rcc} Y(t) - Y(0) & = & Y(t) - Y(T_{N(t)}) + Y(T_{N(t)}) - Y(T_{N(t)}^-) \\ & + & Y(T_{N(t)}^-) - Y(T_{N(t)-1}) + Y(T_{N(t)-1}) - Y(T_{N(t)-1}^-) \\ & + & \cdots \\ & + & Y(T_i) - Y(T_i^-) + Y(T_i^-) - Y(T_{i-1}) \\ & + & \cdots \\ & + & Y(T_1^-) - Y(0), \end{array}$$

that is the sum

$$Y(t) - Y(T_{N(t)}) + \sum_{T_i \leq t} \left[Y(T_i) - Y(T_i^-) + Y(T_i^-) - Y(T_{i-1}) \right].$$

Using the expressions (3.5), (3.6), the characterization of the jump processes and ordering integrals appropriately, we obtain:

$$\begin{array}{rcc} Y(t) - Y(0) & = & \int_0^t \{ \left[u_s(X_s, s) + u_x(X_s, s)\tilde{b}(s) + \frac{1}{2}u_{xx}(X_s, s)\sigma^2(s) \right] \\ & + & \int_E \left[u\left(X_{s^-} + K(s, \xi), s^- \right) - u(X_{s^-}, s^-) \right] \nu(d\xi) \} ds \\ & + & \int_0^t u_x(X_s, s)\sigma(s)dW_s \\ & + & \int_0^t \int_E \left[u\left(X_{s^-} + K(s, \xi), s^- \right) - u(X_{s^-}, s^-) \right] \left(\mathcal{N}(ds, d\xi) - \nu(d\xi)ds \right). \end{array}$$

In the first integral, second line, we are integrating in ds, so it is possibile to replace $u\left(X_{s^-} + K(s, \xi), s^- \right)$ with $u\left(X_s + K(s, \xi), s \right)$. In the last integral instead we use the regularity of u to write $u(s)$

instead of $u(s^-)$. Finally,

$$\begin{aligned} Y(t)-Y(0) &= \int_0^t \{\left[u_s(X_s,s)+u_x(X_s,s)\tilde{b}(s)+\tfrac{1}{2}u_{xx}(X_s,s)\sigma^2(s)\right] \\ &+ \int_E \left[u\left(X(s)+K(s,\xi),s\right)-u(X_s,s)\right]\nu(d\xi)\}ds \\ &+ \int_0^t u_x(X_s,s)\sigma(s)dW_s \\ &+ \int_0^t\int_E \left[u\left(X_{s^-}+K(s,\xi),s\right)-u(X_{s^-},s)\right]\left(\mathcal{N}(ds,d\xi)-\nu(d\xi)ds\right), \end{aligned}$$

that is the target formula, with M_t^u sum of the last two lines. □

Generators

We summarize below the Ito formulas for the process $Y_t = u(X_t, t)$:

- *Formula for diffusive processes in* $\mathbb{R}^n$*:*

$$dY_t = u_t dt + \left(\sum_{i=1}^n u_{x_i}b_i + \frac{1}{2}\sum_{i,j} u_{x_ix_j}\left[\sigma\sigma^t\right]_{ij}\right)dt + \sum_{i=1}^n\sum_{j=1}^m u_{x_i}\sigma^{ij}dW^j,$$

  that, for $n = 1$ is

$$dY = u_t dt + \left(u_x b + \frac{1}{2}u_{xx}\sigma^2\right)dt + u_x\sigma dW;$$

- *Formula for diffusive processes with jumps in* $\mathbb{R}^n$*:*

$$dY = u_t dt + \left(\sum_{i=1}^n u_{x_i}b_i + \frac{1}{2}\sum_{i,j} u_{x_ix_j}\left[\sigma\sigma^t\right]_{ij} + \int_E \left[u(x+K(x,\xi),t)-u(x,t)\right]\nu(d\xi)\right)dt + dM^u,$$

  that, for $n = 1$ is

$$dY = u_t dt + \left(u_x b + \frac{1}{2}u_{xx}\sigma^2 + \int_E \left[u(x+K(x,\xi),t)-u(x,t)\right]\nu(d\xi)\right)dt + dM^u.$$

In each of the previous, the term in brackets, integrated with respect to dt is called *generator*. Generators can be used to obtain a description of the solutions of partial differential equations and have a preponderant role in the next part. So we have:

- *Generator of diffusive processes in* $\mathbb{R}^n$*:*

$$\mathcal{A}u(x,t) = \sum_{i=1}^n u_{x_i}(x,t)b_i(x) + \frac{1}{2}\sum_{i,j} u_{x_ix_j}(x,t)\left[\sigma\sigma^t\right]_{ij}(x),$$

  that, for $n = 1$ is

$$\mathcal{A}u(x,t) = u_x(x,t)b(x) + \frac{1}{2}u_{xx}(x,t)\sigma^2(x);$$

- *Generator of diffusive processes with jumps in* $\mathbb{R}^n$*:*

$$\mathcal{A}u(x,t) = \sum_{i=1}^{n} u_{x_i}(x,t)b_i(x) + \frac{1}{2}\sum_{i,j} u_{x_ix_j}(x,t)\left[\sigma\sigma^t\right]_{ij}(x)$$

$$+ \int_E \left[u(x+K(x,\xi),t) - u(x,t)\right]\nu(d\xi),$$

that, for $n = 1$ is

$$\mathcal{A}u(x,t) = u_x(x,t)b(x) + \frac{1}{2}u_{xx}(x,t)\sigma^2(x) + \int_E \left[u(x+K(x,\xi),t) - u(x,t)\right]\nu(d\xi).$$

Remark. In general, from this point on, we indicate with $\mathcal{A}$ the generator of a diffusion process with jumps, specifying with $\mathcal{L}$ the term exclusively diffusive and with $\mathcal{I}$ the jump term:

$$\mathcal{L}u(x,t) = \sum_{i=1}^{n} u_{x_i}(x,t)b_i(x) + \frac{1}{2}\sum_{i,j} u_{x_ix_j}(x,t)\left[\sigma\sigma^t\right]_{ij}(x),$$

$$\mathcal{I}u(x,t) = \int_E \left[u(x+K(x,\xi),t) - u(x,t)\right]\nu(d\xi).$$

3.2 Kolmogorov backward: Feynman-Kac equation

The Feynman-Kac equation establishes a connection between stochastic differential equations and partial differential equations. A large class of expected values of stochastic processes can be calculated using deterministic methods. Using Ito formula, let X is the stochastic process with generator $\mathcal{A}$, solution of a stochastic differential equation with initial deterministic point $X_0 = x$. Let $\Lambda(x)$ and $h(x)$ be two smooth and limited functions. It is possible to represent the solution of the following problem,

$$\begin{cases} u_t = \mathcal{A}u - \Lambda u, \\ u(t=0) = h. \end{cases}$$

as an appropriate expected value of the process X,

$$u(x,t) = \mathbb{E}\left(h(X(t))\mathrm{e}^{-\int_0^t \Lambda(X(u))du} | X(0) = x\right).$$

This is the Feynman-Kac formula, of which we give a proof in the case of $n = 1$ [75].

Theorem 15. *Let X be the solution of a stochastic differential equation with jumps in the form*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_Z K(X_{t^-},\xi)\mathcal{N}(dt,d\xi),$$

with $X_0 = x$, deterministic, generator $\mathcal{A}$ as in the previous section, and the functions b, σ, K that verify the hypotesis for existence and uniqueness for X. Let $\Lambda(x)$ and $h(x)$ be smooth

and limitated functions (or $h(x)$ limitated and $\Lambda(x)$ non negative). Consider the differential problem

$$\begin{cases} u_t(x,t) = \mathcal{A}u(x,t) - \Lambda(x)u(x,t), & t > 0 \\ u(x,0) = h(x). \end{cases} \tag{3.7}$$

If $u(x,t)$ is a limited and $C^{2,1}$ solution of the problem, it can be represented as

$$u(x,t) = \mathbb{E}\left(h(X_t)e^{-\int_0^t \Lambda(X_u)du}|X_0 = x\right).$$

In particular, this solution is limitated on compacts $[0,T]$ for all $T > 0$ and is unique.

Proof. The sketch of the proof is showing that the process $M = (M_s)_{s\geq 0}$, defined as

$$M_s = u(X_s, t-s)\mathrm{e}^{-\int_0^s \Lambda(X_u)du},$$

is a martingale. If so, since the martingale property $\mathbb{E}(M_t|\mathcal{F}_s) = M_s \, \forall t \geq s$ holds, we can write

$$\mathbb{E}(M_t|X_0 = x) = M_0.$$

We observe that

$$M_0 = u(X_0, t) = u(x,t),$$

so it holds:

$$\begin{aligned} u(x,t) = \mathbb{E}(M_t|X_0 = x) &= \mathbb{E}\left(u(X_t,0)\mathrm{e}^{-\int_0^t \Lambda(X_u)du}|X_0 = x\right) \\ &= \mathbb{E}\left(h(X_t)\mathrm{e}^{-\int_0^t \Lambda(X_u)du}|X_0 = x\right), \end{aligned}$$

the theorem is then proved.

However, it remains to be shown that M is a martingale. To do this we have to calculate dM_s and first, it is useful to apply Ito formula (3.4) to the functione $g(x,s) = u(x,t-s)$ and to the process X obtaining

$$dg(X_s,s) = g_s(X_s,s)ds + \mathcal{A}g(X_s,s)ds + dM_s^g.$$

Considering that

$$\begin{aligned} &g_s(x,s) = -u_t(x,t-s), \\ &g_{xx}(x,s) = u_{xx}(x,t-s), \quad g_x(x,s) = u_x(x,t-s), \end{aligned}$$

and since u is solution of the problem (3.7) we obtain

$$\begin{aligned} du(X_s,t-s) &= \left[-u_t(X_s,t-s) + \mathcal{A}u(X_s,t-s)\right]ds + dM_s^g \\ &= \Lambda(X_s)u(X_s,t-s)ds + dM_s^g. \end{aligned}$$

Before calculating dM we define a new process

$$Y_s = \mathrm{e}^{-\int_0^s \Lambda(X_u)du},$$

that has stochastic differential

$$dY_s = -Y_s\Lambda(X_s)ds.$$

Differential dM can be calculated as:

$$dM_s = d\left[u(X_s, t-s)Y_s\right] = du(X_s, t-s)\cdot Y_s + u(X_s, t-s)\cdot dY_s$$
$$= Y_s\Lambda(X_s)u(X_s, t-s)ds + Y_s dM_s^g - Y_s\Lambda(X_s)u(X_s, t-s)ds,$$

i.e.

$$dM_s = \mathrm{e}^{-\int_0^s \Lambda(X_u)du} dM_s^g,$$

and so

$$M_t = \int_0^t \mathrm{e}^{-\int_0^s \Lambda(X_u)du} g_x(X_s, s)b(X_s)dW_s$$
$$+ \int_0^t \int_E \mathrm{e}^{-\int_0^s \Lambda(X_u)du} \left[g\left(X_{s^-} + K(X_{s^-}, s)\right) - g(X_{s^-}, s)\right]\left(\mathcal{N}(ds, d\xi) - \nu(d\xi)ds\right).$$

The first integral is still a Ito integral, therefore it is a martingale. The second integral is a martingale, since the process

$$\left[g\left(s, X_{s^-} + K(X_{s^-}, \xi)\right) - g(s, X_{s^-})\right],$$

also with the addition of the limited term $\exp\left(-\int_0^t \Lambda(X_u)du\right)$ continues to satisfy the hypotheses. Then M is a martingale and the theorem is proved. □

If in (3.7) we consider the function $\Lambda = 0$ and if it is possible to take h as the indentity function (bounded if, for example, x lives in a bounded interval), then the solution of the problem

$$\begin{cases} u_t(x,t) = \mathcal{A}u(x,t), & t > 0, \\ u(x,0) = x, \end{cases}$$

can be represented as

$$u(x,t) = \mathbb{E}\left(X_t | X_0 = x\right),$$

with this interpretation: *knowing that the process starts from position x, what is its expected position at time t?*

3.3 Kolmogorov forward: Fokker-Planck equation

As said, Feynman-Kac equation answers the question *knowing that the process starts from position x, what is its expected position at time t?*. We could however be interested not so much in the study of the expected position when the initial condition x changes, rather to the evolution of the density of a population that has an initial distribution assigned, and which modifies this distribution over time.

Let X be a random variable on the space $(\Omega, \mathcal{F}, \mathbb{P})$, the law of X is the probability $\mathbb{P}\left(X \in \Gamma\right)$, with $\Gamma \in \mathcal{F}$, region of space on which X is defined. X admits a probability density if a function $\varrho(x)$ exists, such that

$$\mathbb{P}\left(X \in \Gamma\right) = \int_\Gamma \varrho(x)dx.$$

To not specify the region Γ in the definition, we use the compact form

$$\mathbb{P}\left(X \in dx\right) = \varrho(x)dx.$$

Remark. If the random variable is degenerate, i.e. deterministic, $\mathbb{P}(X = x_0) = 1$, there are no density functions. We refer to this eventuality by saying (with abuse of notation) that X admits density δ_{x_0}

Fokker-Planck equation is a partial differential equation that describes the evolution of the probability density of a certain quantity, which is a function of space and time $\varrho(x,t)$, under the influence of forces of deterministic or random origin. A simple and explanatory case is the analysis of the stochastic equation

$$\begin{cases} dX_t = dW_t, \\ X_0 \sim \varrho_0. \end{cases}$$

standard Wiener process, with initial distribution not necessarily deterministic ϱ_0, such that

$$\mathbb{P}(X_0 \in dx) = \varrho_0(x)dx.$$

The Fokker-Planck equation relative to density $\varrho(x,t)$ of this process turns out to be

$$\partial_t \varrho(x,t) = \frac{1}{2}\partial^2_{xx}\varrho(x,t),$$

which has solution, starting from $\varrho_0 = \delta_0$ (degenerate random variable, concentrated in 0)

$$\varrho(x,t) = \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{\frac{-x^2}{2t}},$$

a Gaussian that decrease over time. To get the Fokker-Planck equation, however, it is necessary to give an idea of *Markov's semigroups*. The jump-diffusion processes and the stochastic differential equations fall into the Markov process theory, for which an alternative theory can be studied, with which is possible to give new definitions of generator and Kolmogorov backward equation. What we need to know in this case is just the definition of transition probability, that is

$$\mathbb{P}(X_t \in \Gamma | X_0 = x) = P(t,x,\Gamma),$$

and to know that the processes we have dealt with can be described through a semigroup of operators, i.e. a family of linear operators with properties

$$P_0 = I, \quad P_{t+s} = P_t \circ P_s \quad \text{for } t,s \geq 0.$$

Let $h \in C_b(\mathbb{R}^d)$ be a continuous and limitated function, and define the operator

$$(P_t h) = \mathbb{E}(h(X_t)|X_0 = x) = \int_{\mathbb{R}^d} h(y)P(t,x,dy).$$

This is a linear operator with $P_0 = I$, infact

$$(P_0 h)(x) = \mathbb{E}(h(X_0)|X_0 = x) = h(x).$$

Assuming that $P_t h$ is still a function in $C_b(\mathbb{R}^d)$, the property $(P_{t+s}h)(x) = P_t \circ P_s h(x)$ holds, for the Chapman-Kolmogorov equation ([76]). The semigroup P_t is called Markov semigroup

of the process. By studying the properties of Markov semigroups, properties of process X_t can be obtained. For example, the generator can be obtained in this theory by defining the operator

$$\mathcal{A}h = \lim_{t\to 0} \frac{P_t h - h}{t},$$

the limit is strong, defined in its domain $\mathcal{D}(\mathcal{A})$. The definition implies that, formally, we can write

$$P_t = \mathrm{e}^{\mathcal{A}t},$$
$$\left(\lim_{t\to 0} \frac{\left(\mathrm{e}^{\mathcal{A}t} - I\right)}{t} = \mathcal{A} \lim_{t\to 0} \frac{\left(\mathrm{e}^{\mathcal{A}t} - I\right)}{\mathcal{A}t} = \mathcal{A} \right).$$

Now consider the function $u(x,t) = (P_t h)(x) = \mathbb{E}\left(h(X_t)|X_0 = x\right)$ and calculate its derivative over time,

$$u_t = \frac{d}{dt}(P_t h) = \frac{d}{dt}\left(\mathrm{e}^{\mathcal{A}t} h\right) = \mathcal{A}\left(\mathrm{e}^{\mathcal{A}t} h\right) = \mathcal{A}P_t h = \mathcal{A}u.$$

More, $u(x,0) = P_0 h(x) = h(x)$. As a result, $u(x,t\ )$ satisfies the differential problem

$$\begin{cases} u_t = \mathcal{A}u, \\ u(x,0) = h(x). \end{cases} \tag{3.8}$$

We have therefore formally obtained the Kolmogorov backward or Feynman-Kac equation and we can write

$$u(x,t) = \left(\mathrm{e}^{\mathcal{A}t} h\right)(x),$$

that formally verify (3.8). We can then define the adjoint semigroup P_t^*, that acts on probability measure, having as codomain probability measures,

$$P_t^* \mu(\Gamma) = \int_{\mathbb{R}^d} \mathbb{P}\left(X_t \in \Gamma | X_0 = x\right) d\mu(x) = \int_{\mathbb{R}^d} P(t,x,\Gamma) d\mu(x).$$

Formally, semigroup P_t^* is the adjoint in L^2 of P_t,

$$\int P_t h(x) d\mu(x) = \int h(x) d\left(P_t^* \mu\right)(x).$$

We can then write $P_t^* = \mathrm{e}^{\mathcal{A}^* t}$ with $\mathcal{A}^*$ adjoint in L^2 of $\mathcal{A}$,

$$\int \mathcal{A} f h dx = \int f \mathcal{A}^* h dx. \tag{3.9}$$

With similar reasonings of the case with $\mathcal{A}$, we can obtain the differential equation of Kolmogorov forward, or Fokker-Planck equation, whose unknown variable is the probability density of the process X_t, as stated in the following lemma [76].

Lemma 16. *Let X_t the solution of the equation of Ito $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ with initial condition the random variable X_0, independent of the Brownian motion of the equation, with density $\varrho_0(x)$. Assume that the process X_t has density $\varrho(x,t) \in C^{2,1}(\mathbb{R}^d \times (0,\infty))$. Then ϱ is a solution to the Fokker-Planck equation*

$$\begin{cases} \varrho_t = \mathcal{A}^* \varrho, & \text{for } (x,t) \in \mathbb{R}^d \times (0,\infty) \\ \varrho(t=0) = \varrho_0, & \text{for } x \in \mathbb{R}^d \end{cases} \tag{3.10}$$

where $\mathcal{A}^*$ is the adjoint of the generator of the process, as defined in (3.9).

Proof. Let $\mathbb{E}$ be the expected value with respect to the measure μ with density ϱ_0 of X_0 (in fact, respect to the product measure induced not only by μ, but also by the measure of the Wiener process of the stochastic equation, or the Poisson measure of the jump process, in the presence of jumps). Remembering that with $u(x,t)$ we indicate the solution of the backward equation,

$$\mathbb{E}\left(h(X_t)|\, X_0 = x\right) = u(x,t),$$

$$u(x,0) = h(x),$$

we can obtain the expected value for an initial non-deterministic datum by integrating respect to the density ϱ_0,

$$\mathbb{E}\left(h(X_t)|\, X_0 \sim \varrho_0\right) = \int_{\mathbb{R}^d} u(x,t)\varrho_0(x)dx$$

$$= \int_{\mathbb{R}^d} \left(\mathrm{e}^{\mathcal{A}t}h\right)(x)\varrho_0(x)dx,$$

because it is solution of the backward equation with initial data h

$$= \int_{\mathbb{R}^d} h\left(\mathrm{e}^{\mathcal{A}^*t}\varrho_0\right)(x)dx,$$

by definition of adjoint operator. Since $\varrho(x,t)$ is the density of X_t, we can also write, by definition of expected value, that

$$\mathbb{E}_{\varrho_0}\left(h(X_t)\right) = \int_{\mathbb{R}^d} h(x)\varrho(x,t)dx.$$

By matching the two expressions found for the expected value at time t, we get

$$\int_{\mathbb{R}^d} h\left(\mathrm{e}^{\mathcal{A}^*t}\varrho_0\right)(x)dx = \int_{\mathbb{R}^d} h(x)\varrho(x,t)dx.$$

Using a density argument one we extend the equality to each $u_0 \in L^2(\mathbb{R}^d)$ and deduce that:

$$\varrho(x,t) = \left(\mathrm{e}^{\mathcal{A}^*t}\varrho_0\right)(x),$$

which solves the equation (3.10), differentiating it first, imposing $t = 0$ then. □

The whole theory of Markov semigroups can be made mathematically rigorous, see [76]. We want here to undestand the form of the operator $\mathcal{A}^*$, which we can obtain explicitly by calculating the adjoint of the generator $\mathcal{A}$. Recalling therefore that for a process of pure diffusion the generator is

$$\mathcal{L}u(x,t) = \sum_{i=1}^{n} u_{x_i}(x,t)b_i(x) + \frac{1}{2}\sum_{i,j} u_{x_ix_j}(x,t)\left[\sigma\sigma^t\right]_{ij}(x),$$

we can calculate $\mathcal{L}^*$ in the case of the real line so that for each u, ϱ in $C_b(\mathbb{R})$ holds (ignoring the time dependency of functions)

$$\int_{\mathbb{R}^n} \left(\mathcal{L}u\right)(x)\varrho(x)dx = \int_{\mathbb{R}^n} u(x)\left(\mathcal{L}^*\varrho\right)(x)dx.$$

Integrating twice

$$\int_{\mathbb{R}} (\mathcal{L}u)(x)\varrho(x)dx = \int_{\mathbb{R}} \left(u_x(x)b(x) + \frac{1}{2}u_{xx}(x)\sigma^2(x) \right) \varrho(x)dx,$$

and canceling the boundary contributions we obtain

$$= \int_{\mathbb{R}} u(x) \left((-b(x)\varrho(x))_x + \frac{1}{2}(\sigma^2(x)\varrho(x))_{xx} \right) dx = \int_{\mathbb{R}} u(x) \left(\mathcal{L}^*\varrho\right)(x)dx,$$

i.e.

$$\mathcal{L}^*\varrho(x) = (-b(x)\varrho(x))_x + \frac{1}{2}(\sigma^2(x)\varrho(x))_{xx},$$

and its generalization in $\mathbb{R}^n$

$$\mathcal{L}^*\varrho(x) = -\sum_i \frac{\partial}{\partial x_i}(b_i(x)\varrho(x)) + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \left(\left[\sigma\sigma^t\right]_{i.j}(x)\varrho(x) \right). \tag{3.11}$$

Remark. If the process is a diffusion with jumps, i.e, X_t is solution of $dX_t = b(X_t)dt + \sigma(X_t)dW_t + J_t$ the theorem remains valid, but the generator $\mathcal{A}$ of the complete process should be considered, with the addition of the jump term

$$\mathcal{I}u(x,t) = \int_E \left[u(x + K(x,\xi),t) - u(x,t)\right] \nu(d\xi).$$

We obtain the explicit form of this operator in the specific case of the model presented in the next chapter.

Part II

Evolutionary dynamics with rare mutations

Chapter 4

Rare mutations

4.1 Random mutations replicator dynamics

We show in this section, using all the theory developed in the previous sections, a model of stochastic differential equations that describe with a greater detail than the classical replicator dynamics systems of the evolving populations. The model, presented in [4], extends the replicator mutator model to the case of random mutations. The original contributions presented in the following two chapters will use this model as a basis from which to start, reformulating it and extending it first to the case of a heterogeneous environment (Chapter 5), then to a population of infinite genotypes (Chapter 6).

Quasispecies

The replicator dynamics completely ignore rare mutations. Unfortunately ignoring them means ignoring one of the driving forces of Evolution, so it is necessary to modify the previous models by adding new terms. Manfred Eigen and Peter Schuster formulated the quasispecies theory [85], A quasispecies is an group of similar genomic sequences generated by a mutation-selection process. In chemistry the word "species" refers to a group of identical molecules, but the species of all RNA molecules does not contain identical sequences, so it is necessary to introduce the concept of "quasispecies". During replication of a genome, mistakes can happen; the probability that replication of individuals of type i results in individuals of type j is given by q_{ij}. This define $Q = (q_{ij})$, that is a stochastic matrix; each rows sums to one, and each $q_{ij} \geq 0$. The quasispecies equation is then defined as:

$$\begin{cases} \dot{x}_k = \sum_{i=1}^{n} x_i f_i q_{ik} - \bar{f} x_k \\ k = 1, \ldots, n \end{cases}$$

The variation of fraction of individuals k is obtained by individuals of type i that mutate, at rate f_j times the probability that replication j generates type i. The sum of x_k remains constant because of the presence of $\bar{f}$. Note that, when the replication is error-free, i.e. there are no mutations, Q is the identity matrix and the quasispecies equation reduces to Replicator Dynamics (1.2) with constant fitness.

Consider as initial condition $x(0)$ in the interior of the simplex, i.e. $x_i(0) > 0$ for all i; then the replicator equation converges to a homogeneous population that consists only of

individuals of the fittest type. If $f_0 > f_i$ for all $i \neq 0$, then the stable equilibrium is given by $x_0 = 1$ and $x_i = 0$.

When the matrix Q is not the identity, mutations occur. This means that there exists $q_{ij} > 0$ with $i \neq j$. In addition, we assume that $f_i > 0$ for at least one i. In this case, the equation of the quasispecies admits a single equilibrium x^*, that is globally stable, and does not necessarily maximize the average fitness $\bar{f}$ ([73, 24]). Consider again $f_0 > f_i$ for all $i \neq 0$. The population of individuals of the type 0 at the equilibrium will have higher fitness than the population at equilibrium. Mutations reduce the average fitness at equilibrium.

The quasispecies equation can be reformulated, giving more emphasis to the similarity with the replicator equation, as the *replicator mutator* (see [73, 24, 10]) :

$$\begin{cases} \dot{x}_k = \left(f_k(x) - \bar{f}(x)\right) x_k + \sum_{i=1}^{n} f_i(x) m_{ik} x_i, \\ k = 1, \ldots, n. \end{cases} \tag{4.1}$$

The first addend is nothing else than the replicator dynamics for a population of n distinct types $x = (x_1, \ldots, x_n)$ with non constant fitness, to which is added a term given by the matrix $M = (m_{ik})_{i,k=1,\ldots,n}$, effective mutations matrix, which is null in the absence mutations and has the form $M = Q - I$, with Q as above. The new term $\sum_{i \neq k} f_i(x) m_{ik} x_i \geq 0$ describes the increase in frequency of individuals of type k due to the birth of mutants from other individuals, while the term $f_k(x) m_{kk} x_k \leq 0$ measures the decrease in the frequency of individuals of type k caused by the presence of mutated descendants among the progeny of such individuals. In this model we are implicitly assuming that mutations occur homogeneously: at each generation, a fixed proportion of the progeny will show mutant traits. We want to propose a description where mutations should not be considered deterministic terms, rather we would like to mimic random changes of the genome, rare "jumps" that randomly modify the frequencies in the population.

Random mutation replicator mutator.

Following [4], we transform (4.1) into a stochastic equation imposing the condition that the mutations occur at random times and are therefore described by a jump process. So be $(T_n, Z_n)_n$ a succession of times T and pairs Z, random on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. The pairs Z_n are chosen in the space

$$Z = \left\{(i,k) \in \{1, \ldots, N\}^2,\, i \neq k\right\},$$

therefore we are assuming that each mutation has a fixed ancestor and a descendant of a single different type. Let

$$N_t = \sum_n \mathbb{I}(T_n \leq t), \quad N_t^{ik} = \sum_n \mathbb{I}(Z_n = (i,k)) \mathbb{I}(T_n \leq t),$$

be the counting processes that evaluate respectively the total number of mutations and the number of mutations from type i to the type k, so that $N_t = \sum_{i \neq k} N_t^{ik}$. Suppose then that the proportion between the descendants of individuals of type i showing the type k after a mutation is constant and equal to $\gamma_{ik} \in (0, 1]$. The stochastic differential equation obtained with these premises is:

$$dX_t^k = a_k(X_t)dt + \sum_{i \neq k} \gamma_{ik} X_t^i dN_t^{ik} - \sum_{i \neq k} \gamma_{ki} X_t^k dN_t^{ki}. \tag{4.2}$$

The function $a_k(X_t) = \left(f_k(X_t) - \bar{f}(X_t)\right) X_t^k$ is the term of the replicator dynamics with $k = 1, \ldots, N$. The initial datum can be deterministic or random with initial distribution ϱ_0 with support in the simplex $\mathcal{S}^N$. We can write the equation in integral form and make the jump process explicit,

$$X_t = X_0 + \int_0^t a(X_s)ds + \int_0^t \int_Z K(X_{s^-}, z)\mathcal{N}(ds, dz). \tag{4.3}$$

The Poisson measure $\mathcal{N}(dt, dz)$ is such that, for all i and k distinct, the intensity of the process, that is the frequency of the jumps, depends on the genetic distance between the type i and the type k but also from the selection: the bigger is the fitness $f_i(x_{t^-})$, the more types i will reproduce, more often the offspring will undergo mutations; for this reason the jumps have intensity $\lambda_{ik} f_i(X_t^-)$.

However, when the fitness vector is constant on the simplex, then the (4.3) can be written with $K\left(x, (i,k)\right) = \mathbb{I}_S(x)\gamma_{ik}x_i(e_k - e_i)$ and with $\mathcal{N}\left([0,T] \times (i,k)\right) = N_t^{ik}$, that is a simple Poisson process with intensity $\nu(\{(i,k)\}) = \lambda_{ik} f_i$. The same construction can be obtained by considering the case of non-constant fitness, linked to the payoff matrix of a given game, but with intensity of the process independent of fitness, with $\nu(\{(i,k)\}) = \lambda_{ik}$.

Existence and uniqueness of the frequency process (4.2) is proved in [4], where the regularity of f ensures that not only the field a verify all the hypothesis of lipschitzianity and sublinearity necessary for the existence and uniqueness of the solution, but also that the intensity of the jump process is bounded. It can then be noted that, as in the case of the replicator dynamics, the simplex is an invariant region for the dynamics of (4.2). If the initial data distribution support X_0 is in $\mathcal{S}^N$ then X_t remains in the simplex for every positive time. It is easy to note that when $x \in \mathcal{S}^N$ then $x + \gamma_{ik}x_i(e_k - e_i) \in \mathcal{S}^N$.

Remark. In the simple case in which the jump process N_t^{ik} is Poisson of constant intensity $\lambda_{ik} f_i$ then the equation (4.2) can be reformulated as:

$$\begin{aligned} dX_t^k = &\left[\left(f_k(X_t) - \bar{f}(X_t)\right) X_t^k + \sum_{i=1}^N f_i \lambda_{ik}\gamma_{ik} X_t^i\right] dt \\ &+ \sum_{i \neq k} \gamma_{ik} X_{t^-}^i dM_t^{ik} - \sum_{i \neq k} \gamma_{ki} X_{t^-}^k dM_t^{ki}, \end{aligned}$$

where, compared to the generic case, the function $\gamma = \gamma_{ik} X_t^i$, the intensity of the process $\lambda = f_i \lambda_{ik}$, the martingale $dM = dM^{ik}$. If we take $m_{i,k} = \lambda_{ik}\gamma_{ik}$ for $i \neq k$, $m_{k,k} = -\sum \lambda_{ki}\gamma_{ki}$, then the equation (4.2) is the *replicator dynamics* perturbed by random terms of martingale type. The relationship

$$m_{ik} = \lambda_{ik}\gamma_{ik},$$

creates a parallelism between the deterministic and the random model, highlighting that there are more pairs $(\lambda_{ik}, \gamma_{ik})$ which give rise to the same replicator mutator model. When m_{ik} is null, we can suppose that also γ_{ik} is null and both models are reduced to the replicator dynamics. In other cases, we can suppose γ_{ik} as free parameter and consequently $\lambda_{ik} = m_{ik}/\gamma_{ik}$. When γ_{ik} tends to 0, the temporal intensity λ_{ik} grows and the trajectories of the process become continuous. On the contrary, when $\gamma_{ik} = 1$, the temporal intensity reaches its minimum in m_{ik} and mutations are concentrated in rare events that occur simultaneously throughout the offspring.

4.2 Expected values and qualitative study

We can obtain Feynmann-Kac equations for the previous model. Consider the equation (4.2). Following the definition of generator we obtain the equation:

$$\begin{cases} \partial_t u_k(x,t) = a(x) \cdot \nabla u_k(x,t) + \mathcal{I} u_k(x,t), & x \in \mathcal{S}^n,\ t > 0 \\ u_k(x,0) = x_k, & x \in \mathcal{S}^n, \end{cases} \tag{4.4}$$

with

$$a(x) = (a_1(x), \ldots, a_n(x)),$$
$$a_k(x) = \left(f_k(x) - \bar{f}(x)\right) x_k,$$

and with the generator of the jump process $\mathcal{I}$,

$$\mathcal{I}u(x,t) = \sum_{i \neq j} \lambda_{ij} f_i(x) \left[u(x + \gamma_{ij} x_i (e_j - e_i), t) - u(x,t)\right].$$

We have chosen, with respect to the hypotheses of the Feynman-Kac theorem, $\Lambda = 0$ and $h = id$ (because it is limited to $\mathcal{S}^n$), then the unknown variable u_k is the expected value of the $k-$th component of x.

Analysis of two-species case.

We now dedicate to the analytical study of the equation, to get an idea of the behavior of the expected frequencies, in the simplest case: two species with constant fitness. We are therefore considering the vector field of the differential equation

$$\dot{x} = -sx(1-x),$$

with $s = f_0 - f_1$ e $x \in [0,1]$, and reducing the number of variables

$$\begin{bmatrix} x_0 = 1 - x \\ x_1 = x \end{bmatrix}.$$

If we consider the constant $s > 0$ we have then that the only asymptotically stable point of equilibrium is $x_0 = 1$, $x_1 = 0$, that is, $x = 0$. The Feynman-Kac equation obtained from (4.4), for the function

$$u(x,t) = \mathbb{E}\left(x(t) | x(0) = x\right),$$

is

$$\begin{cases} \partial_t u + s(1-x)x\partial_x u = \lambda_0 f_0 \mathcal{I}_0 u + \lambda_1 \mathcal{I}_1 u, & 0 \leq x \leq 1, t > 0 \\ u(x,0) = x, & 0 \leq x \leq 1 \end{cases} \tag{4.5}$$

where

$$\mathcal{I}_0 u(x,t) = u(x + \gamma_0(1-x), t) - u(x,t),$$
$$\mathcal{I}_1 u(x,t) = u(x - \gamma_1 x, t) - u(x.t).$$

Remark. It can be seen that, in the absence of the mutation term, the (4.5) is the equation of homogeneous transport that came from replicator dynamics.

We abbreviated the notations by writing $\gamma_0 = \gamma_{01}$, $\gamma_1 = \gamma_{10}$, $\lambda_0 = \lambda_{01}$, $\lambda_1 = \lambda_{10}$. In this simple case we can analytically compare the dynamics of the stochastic model and the traditional one of replicator mutator, which in this context has the form

$$\begin{cases} \dot{x} = -sx(1-x) + m_0 f_0(1-x) - m_1 f_1 x, & t > 0 \\ x(0) = x, \end{cases}$$

with m_0 parameter for mutation from species 0 to species 1, m_1 vice versa. As we noted, we can take, for $i = 0, 1$,

$$m_i = \lambda_i \gamma_i,$$

and, according to the previous observation, we consider the equation of homogeneous transport linked to the replicator mutator, with unknown variable $X(x,t)$,

$$\begin{cases} \partial_t X + (sx(1-x) - m_0 f_0(1-x) + m_1 f_1 x)\partial_x X = 0, & 0 \le x \le 1, t > 0 \\ X(x,0) = x. & 0 \le x \le 1 \end{cases} \tag{4.6}$$

Now [4] present a series of results,

Proposition 17. *The following results on the regularity of the solutions hold:*

- *The problem (4.5) admit a classic solution $u \in C^\infty([0,1] \times [0,\infty))$;*
- *For all $t > 0$ and for $x \in [0,1]$, the solution of (4.5) satisfyies*

$$0 \le \partial_x u(x,t) \le e^{(s - m_0 f_0 - m_1 f_1)t};$$

- *For each $t > 0$ and for $x \in [0,1]$, the solution of (4.5) is convex and there exist two constants $c > 0$ and $\mu \in \mathbb{R}$ such that*

$$0 \le \partial^2_{xx} u(x,t) \le c e^{\mu t}.$$

Thanks to this, it is possible to understand that the expected value of the population quantity of the stochastic equation is greater than or equal to that of the deterministic case, therefore rare mutations increase the survival opportunities of the lower-fitness species. This is the result contained in the following proposition,

Proposition 18. *Let u and X be respectively solution of (4.5) and (4.6). Then*

$$1 \ge u(x,t) \ge X(x,t),$$

for all $0 \le x \le 1$ and $t \ge 0$.

Proof. The solution of (4.5) is $u(x,t) \le 1$ because the constant 1 is a supersolution. We show that $u(x,t)$ is a supersolution of (4.6):

$$\partial_t u + (sx(1-x) - m_0 f_0(1-x) + m_1 f_1 x)\partial_x u =$$

$$\partial_t u + sx(1-x)\partial_x u - m_0 f_0(1-x)\partial_x u + m_1 f_1 x \partial_x u =$$

$$\lambda_0 f_0 \mathcal{I}_0 u + \lambda_1 \mathcal{I}_1 u - m_0 f_0 (1-x)\partial_x u + m_1 f_1 x \partial_x u =$$
$$\lambda_0 f_0 \left[u(x+\gamma_0(1-x),t) - u(x,t) - \gamma_0(1-x)\partial_x u(x,t)\right]$$
$$+\lambda_1 f_1 \left[u(x-\gamma_1 x,t) - u(x,t) - \gamma_1 x \partial_x u(x,t)\right] =$$
$$\lambda_0 f_0 \int_0^1 \partial_{xx} u(x+\theta\gamma_0(1-x),t)d\theta + \lambda_1 f_1 \int_0^1 \partial_{xx} u(x-\theta\gamma_1 x,t)d\theta \geq 0,$$

for the convexity. □

The analysis addressed in [4] goes further, trying to show if, under appropriate parameter choices, the dynamics of random mutations can provide unexpected results with respect to the classical replicator mutator. First of all we can state a lemma about the existence of solutions for long times.

Proposition 19. *For each choice of the parameters, the function*

$$\bar{u}(x) = \lim_{t\to\infty} u(x,t),$$

is well defined for all $x \in [0,1]$*. If, then,* $m_1 > 0$*, or* $m_1 = 0$ *and* $m_0 \geq s/f_0$*, then the limit* $\bar{u}$ *is constant.*

For some parameter choices, the random model gives the same results as the classic problem. This happens, for example, if $m_0 = 0$, when there are no mutations of type 0 to type 1. In this case the mutated descendants have higher fitness than the progenitors, and mutations help selection in fixing type 0 in the population. Even in the opposite situation, when $m_1 = 0$ and $m_0 \geq s/f_0$, the behavior of the two models coincides. In this case the only mutations are those from the highest fitness type, and the mutation rate is high enough to overcome the mutation effect. In the case where instead the coefficient m_1 is null, but the mutation rate m_0 is not sufficiently high, i.e. $m_0 < s/f_0$, something new happens. The final state depends on the intensity of the jump process that governs the mutations and no longer necessarily follows the deterministic model of replicator mutator. All these statements are mathematically justified in the aforementioned article.

Chapter 5

Spatial heterogeneous environment

We propose in this chapter a stochastic model that is an extension to the spatial case of [4], where individuals can mutate changing their strategies randomly (but rarely) and explore the external environment. This environment affects the selective pressure by modifying the payoff arising from the interactions between strategies. We derive a Fokker-Plank integro-differential equation and provide Monte Carlo simulations for the Hawks vs Doves game. In particular we show that, in some cases, taking into account the external environment favors the persistence of the low-fitness strategy. This chapter is the extended version of the article [5].

5.1 Introduction

Evolutionary Dynamics describes biological systems subject to Darwinian Evolution by taking into account the main mechanisms and phenomena of Evolution itself. In [60], Maynard Smith and Price propose an instance of this approach by considering a population modified according to the replicator dynamics. A population is formed by d *types*, or *behaviors*, $E_1, \ldots, E_d$, with fractions corresponding to relative abundance in the vector $x = (x_1, \ldots, x_d)$, which corresponds to a point in the simplex $\mathcal{S}^d$. The selection and adaptation mechanism is described by means of a system of differential equations in the following form:

$$\frac{\dot{x}_k}{x_k} = f_k(x) - \bar{f}(x), \tag{5.1}$$

as $k = 1, \ldots d$. The rate of increment $\dot{x}_k/x_k$ of the type E_k is given by its absolute fitness, denoted with f_k, balanced with the average fitness of the population $\bar{f}$, which has the form

$$\bar{f}(x) = \sum_{k=1}^{d} x_k f_k(x).$$

In evolutionary matrix game theory the vector of absolute fitness $f = (f_1, \ldots, f_d)$ is defined as

$$f(x) = \mathcal{U}(x)\, x,$$

where $\mathcal{U}(x)$ is the matrix of payoff that rules the interplay between different strategists (and possibly depends on the frequencies of different species themselves). In this regard, the fitness

of the type E_k is defined as the result that an individual of that type gets colliding against another individual on average, i.e.

$$f_k(x) = [\mathcal{U}(x)\, x]_k = \sum_{i=1}^{d} u_{ki}(x)\, x_i.$$

However, it is clear that the basic element for the generation of evolutionary novelties are mutations. The quasispecies equation, dating back to the 1970s, modifies the growth rate of each species by considering the dispersion due to the birth of mutated offspring [85]:

$$\dot{x}_k = \sum_{i=1}^{d} f_i\, q_{ik}\, x_i - \bar{f} x_k. \tag{5.2}$$

Here the coefficient q_{ik} express the proportion of offspring of k-type from a progenitor i, which shows up at any procreation. An important aspect of mutations stands in their randomness, which is quite underrated in (5.2). Since then many more refined models have been proposed to put into the right light randomness; we refer for instance to [14] showing that one single stochastic microscopic process can generate different macroscopic models of adaptive evolution. More recently, in [4], it has been proposed a macroscopic stochastic model where mutations occur at a different time scale than selection. This approach goes into the direction of adaptive dynamics, but differentiates from trait substitution sequence because it is not assumed that there is complete adaptation (namely invasion or extinction of the mutant trait) between subsequent mutations. Within the framework of social dilemma, where the types E_i are read as strategies, a "mutation" happens when a player changes his strategy. The model in [4] assumes that such events happen on rare and random occasions, even more than once before the system reaches its stable state. See the previous chapter for a more detailed review of [4], and see also the numerical paper [3], focused on Prisoner's Dilemma.

In this Chapter we take a step further and address our attention to the environment, seen as a place where individuals can evolve but also as a factor that can influence the dynamics of interaction between strategists. The model presented in [4] is then expanded to take into account how the natural environment can modify the interactions between individuals, changing selective pressures. We add a new variable $y \in \mathbb{R}^N$ to the variable x, in the simplex, so that the status of the population is described by the pair (x, y). The new variable y stands for the position of the population or, more widely, for an external parameter that affects the results of the interplay between strategies. It changes according to a velocity, partly deterministic, partly stochastic, and influences the selection mechanism because the payoff matrix depends on y. In the following Section 5.2 we recall the stochastic model for replicator dynamics with point-type mutations introduced in [4]. With the aim of performing Monte-Carlo simulations, we give an alternative (but equivalent) description of the process by using a single Poisson random measure. Starting from this description, we generate an algorithm to simulate our process. Next, the spatial environment is introduced as a further stochastic variable, whose dynamics is ruled by a SDE. Therefore, we end up with two coupled SDE for the character-position variables (x, y): see (5.7), (5.8).

In Section 5.3 we derive a Fokker-Plank integro-differential equation for (5.7), (5.8), (see (5.12) later on). The classical regularity assumptions requested by the Hormander theory are not satisfied because of the presence of a non-local term, which is the deterministic counterpart

of the point process modeling mutations. We therefore read it in the viscosity sense, even if the problem (5.12) does not fit plainly in the standard framework of viscosity solutions for integro-differential equations: the main difficulty comes from the domain where it is set, which is closed. Actually, the model does not justify any attempt to impose a boundary condition. Moreover the nonlocal term does not depend continuously on x. These difficulties are overcome by extending in a suitable way the problem to the whole space (5.13) and noticing that the produced solution can actually be interpreted as a probability density for the couple character-position (x, y).

Finally Section 5.4 provides numerical simulations concerning the two strategist game Hawks vs Doves, used by Maynard Smith to explain the high frequency of conventional displays, rather than all-out fight, among animals (especially within heavily armed species) [44]. We modify the standard model by assuming that the cost for fighting changes according to the location, and perform various simulations for the probability density obtained both by a Monte-Carlo method starting from the stochastic system (5.7), (5.8), and by a finite difference scheme based on the Fokker-Plank equation (5.13). The equilibrium of the standard replicator-mutator dynamics can be disrupted by effect of either random motion or mutations. In some particular cases, the environment itself allows for the survival of the low fitness species.

5.2 A stochastic model for mutations in heterogeneous environment

We propose to describe the frequencies of different phenotypes in the population according to a stochastic differential equations (SDE) in the general mathematical framework (see Chapter 2):

$$X_t = X_0 + \int_0^t a\left(X_s\right) ds + \int_0^t b(X_s) dW(s) + \int_0^t \int_E K(X_{s^-}, \xi) \mathcal{N}(ds, d\xi). \tag{5.3}$$

Here X_t is a process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where a, b, K are Borel measurable functions of appropriate dimensions. $W(s)$ is a standard Brownian motion and $\mathcal{N}(ds, d\xi)$ is a Poisson random measure on $\mathbb{R}^+ \times E$, with mean measure $l \times \nu$, l Lebesgue measure on $\mathbb{R}^+$, ν a σ-finite measure on a measurable space $(E, \mathcal{E})$.

The process of classic replicator dynamics (5.1) is obtained when $X = (x_1, \ldots x_d)$ is the vector of relative frequencies of d various phenotypes, a is the vector of relative fitness, i.e. $a(X) = (\ldots, a_k(X), \ldots)$, with

$$a_k(X) = x_k \left(f_k(X) - \bar{f}(X)\right),$$

and b and K are null, so that (5.3) is totally deterministic.

In [4], mutations are described by means of a pure point process that alters replicator dynamics and the Brownian motion term is zero ($b = 0$). Any mutation has a fixed progenitor (type i) and a unique descendant (type j): this gives $2\binom{d}{2} = d(d-1)$ different mutations, precisely all those that transform a type i in a type j as

$$(i, j) \in I = \left\{(i, j) \in \{1, \ldots, d\}^2 ; i \neq j\right\}.$$

The mutation from type i to type j is driven by a non-homogeneous point process N_t^{ij} with stochastic intensity $\lambda_{ij} f_i(X_{t^-})$. The process N_t^{ij} makes unit jumps with a frequency depending on the process itself, according to the "genetic distance" between the types i and j (λ_{ij}) and the fitness of i (f_i): the higher the fitness, the higher the rate of reproduction of individuals of that kind, the more they will suffer mutations. A further coefficient $\gamma_{ij} \in (0,1)$ measures the proportion of individuals involved in mutations: the population of type i decreases by a fraction $\gamma_{ij} x_i$, while the population of type j increases by the same amount. This yields a jump of the population frequency vector of size $\gamma_{ij} x_i (e_j - e_i)$, e_i standing for the unit vector pointing in the direction i. The resulting SDE is

$$x_{k,t} = x_k(0) + \int_0^t a_k(X_s) ds + \sum_{i \neq k} \int_0^t \gamma_{ik} x_{i,t} dN_t^{ik} - \sum_{i \neq k} \int_0^t \gamma_{ki} x_{k,t} dN_t^{ki}. \tag{5.4}$$

Let us notice by now that the number of variables depicting the character can be reduced by observing that $x_d = 1 - \sum_{i=1}^{d-1} x_i$ and setting the problem in the closed set

$$\Sigma^d = \{(x_1, \dots x_{d-1}) \,:\, x_i \geq 0, \sum_{i=1}^{d-1} x_i \leq 1\}.$$

With a little abuse of notations we shall continue to write $x \in \Sigma^d$ and

$$\begin{aligned} f_k(x) =& f_k(x_1, \dots x_{d-1}, 1 - \sum_{i=1}^{d-1} x_i), \\ a_k(x) =& a_k(x_1, \dots x_{d-1}, 1 - \sum_{i=1}^{d-1} x_i) = (f_k - f_d)(1 - x_k) x_k - \sum_{\substack{i=1 \\ i \neq k}}^{d-1} (f_i - f_d) x_i x_k. \end{aligned}$$

In the same paper [4], a Kolmogorov integro-differential equation describing the expected frequencies is derived and investigated analytically, with particular attention to the long term equilibrium. Analytical investigation is satisfactory in the case of constant fitness (quasispecies equation), but there are some gaps concerning variable fitness, that has been tackled by a numeric approach in the subsequent paper [3]. In the present work we are mainly concerned with Monte-Carlo simulations. That is why, before enriching the model by including the effect of heterogeneous environment, it is worth giving an alternative description and present an algorithmic approach.

The SDE (5.4) can be written in standard form (5.3) by taking $d(d-1)$ independent Poisson random measures $\mathcal{N}_{ij}(ds, d\xi)$ on $\mathbb{R}^+ \times \mathbb{R}^+$, defining the amplitudes of jumps as

$$K_{ij}(X, \xi) = \gamma_{ij} x_i (e_j - e_i) 1_{[0, \lambda_{ij} f_i(X))}(\xi), \tag{5.5}$$

and then invoking the Poisson embedding [16].

It is possible to set up an equivalent mode (i.e. with the same probability distribution) with only one random measure $\mathcal{N}(ds, d\xi)$ on $\mathbb{R}^+ \times E$ with $E = \mathbb{R}^+ \times [0,1]$. To this aim we look at the sum of the stochastic intensity of each individual process

$$\Lambda(X) = \sum_{i \neq j} \lambda_{ij} f_i(X),$$

split the unit interval into $d(d-1)$ disjoint intervals $\mathcal{I}_{ij}$ of length $\lambda_{ij} f_i(X)/\Lambda(X)$, and take the amplitude of jumps as

$$K(X, \xi) = K(X, u, \theta) = 1_{[0,\Lambda(X))}(\theta) \sum_{i \neq j} \left[\gamma_{ij} x_i (e_j - e_i) 1_{\mathcal{I}_{ij}}(u) \right]. \tag{5.6}$$

The two processes just described coincide indeed.

Lemma 20. *The processes (5.5) and (5.6) have the same infinitesimal generator, so they have the same probability distribution.*

Proof. The generator of (5.6) is

$$\begin{aligned}
&\int_{(0,1)} \int_{\mathbb{R}} \left[\phi\left(X + 1_{(0,\Lambda(X)]}(\theta) \textstyle\sum_{i \neq j} \gamma_{ij} x_i (e_j - e_i) 1_{\mathcal{I}_{ij}}(u) \right) - \phi(X) \right] du d\theta \\
&\quad = \Lambda(X) \int_{(0,1)} \left[\phi\left(X + \textstyle\sum_{i \neq j} \gamma_{ij} x_i (e_j - e_i) 1_{\mathcal{I}_{ij}}(u) \right) - \phi(X) \right] du \\
&\quad\quad = \Lambda(X) \textstyle\sum_{i \neq j} \int_{\mathcal{I}_{ij}} \left[\phi\left(X + \gamma_{ij} x_i (e_j - e_i) \right) - \phi(X) \right] du \\
&\quad\quad\quad = \Lambda(X) \textstyle\sum_{i \neq j} |\mathcal{I}_{ij}| \left[\phi\left(X + \gamma_{ij} x_i (e_j - e_i) \right) - \phi(X) \right] \\
&\quad\quad\quad\quad = \textstyle\sum_{i \neq j} \lambda_{ij} f_i(X) \left[\phi\left(X + \gamma_{ij} x_i (e_j - e_i) \right) - \phi(X) \right],
\end{aligned}$$

i.e. the same infinitesimal generator of (5.5), as in [4]. □

This alternative construction, albeit equivalent to the first one, can be turned into a simulation more easily and with a more compact and efficient code, because it involves only one jump process instead of $d(d-1)$ independent ones. In view of Monte Carlo approximations, we therefore give an intuitive interpretation of this last process, based on the existence theorem for Poisson random measures in [50].

Let $T > 0$ a fixed time horizon and

$$\Lambda^{\max} = \max_X \Lambda(X).$$

The evolution process can be simulated by the following steps:

1. Build a priori a homogeneous Poisson process with intensity $\Lambda^{\max}$, whose jump times will be denoted by T_n lower than T;
2. Simulate the replicator dynamics till T_1;
3. Extract uniformly a random number $\xi \in [0, 1]$;
   (a) if $\Lambda^{\max} \xi > \Lambda(X_{T_1^-})$ no jump occurs,
   (b) if $\Lambda^{\max} \xi \leq \Lambda(X_{T_1^-})$ a jump occurs indeed. To decide which kind of mutation occurs, extract another random number $u \sim \text{Unif}(0, 1)$ and look at which interval $\mathcal{I}_{\hat{i}\hat{j}}$ it belongs (which is always the case because the sets $\mathcal{I}_{ij}$ form a partition of $[0, 1]$). Then shift a quantity $\gamma_{\hat{i}\hat{j}} x_{\hat{i}, T_1^-}$ from $\hat{i}$ to $\hat{j}$.
4. Restart from step 2.

5.2.1 Heterogeneous environment

In the present model the only observed variables are the frequencies of the various phenotypes, as well as in the classical replicator equation. The rules of the play are fixed once and for all by means of the payoff matrix $\mathcal{U}$, and nothing depends on the physical position of the population, as if the individuals were not able to move, or if the environment were completely homogeneous. A more realistic picture has to take into account that environmental changes affect the results of interaction between different behaviors.

To introduce heterogeneous environment we increase the observed variables so that the status of the population is described by a pair $X = (x, y)$: as before $x = (x_1, \dots, x_{d-1}) \in \Sigma^d$ stands for the *character* of the population, each x_i being the fraction of individuals of type E_i (and $x_d = 1 - \sum_{i=1}^{d-1} x_i$ the fraction of type E_d), while the new variable $y \in \mathbb{R}^N$ stands for the *position* of the population. More widely this new variable can be seen as an external parameter that affects the results of the interplay between strategies. The payoff matrix depends on y, i.e. $\mathcal{U} = \mathcal{U}(y)$, consequently also the respective fitness

$$f_k(x, y) = \sum_{i=1}^{d-1} u_{ki}(y)x_i + u_{id}(y)(1 - \sum_{i=1}^{d-1} x_i)$$

varies with y.

The character x evolves according to a suitable version of equation (5.4):

$$x_t = x_0 + \int_0^t a\,(x_s, y_s)\,ds + \int_0^t \int_E K(x_{s^-}, y_s, \xi)\mathcal{N}(ds, d\xi). \tag{5.7}$$

Here

- $a \in \mathbb{R}^{d-1}$ stands for the *vector field of the replicator dynamics*. It has the same structure as in the former case, but with an important difference: the fitness are allowed to depend from y, so that

$$a_k(x, y) = x_k(f(x, y) - \bar{f}(x, y)) \quad \text{as } k = 1, \dots d.$$

- The jump amplitude K and the random measure $\mathcal{N}$ describe the *mutation process* as before. The location y affects the mutation process through the fitness, as

$$\Lambda(x, y) = \sum_{i \neq j} \lambda_{ij} f_i(x, y),$$

$$K\,(x, y, u, \theta) = 1_{[0,\Lambda(x,y))}(\theta) \sum_{i \neq j} \gamma_{ij} x_i\,(e_j - e_i)\,1_{\mathcal{I}_{ij}(x,y)}(u),$$

  where the intervals $\mathcal{I}_{ij}(x, y)$ have length equal to $\lambda_{ij} f_i(x, y)/\Lambda(x, y)$ and form a partition of the unit interval, as $i \neq j \in \{1, \dots, d\}$.

The environmental variable y changes according to a diffusion with drift:

$$y_t = y_0 + \int_0^t v\,(x_s, y_s)\,ds + \int_0^t \sigma\,(x_s, y_s)\,dW_s, \tag{5.8}$$

where

- $v \in \mathbb{R}^N$ stands for the *velocity field of the population*. For any given y, $v(e_i, y)$ is the drift of the type E_i, while a composite population described by the character x is inclined to move according to $v(x, y)$.
- σ is an $N \times N$ matrix and W_s is an N-dimensional Brownian motion, describing the random component of the displacement.

Notice that both the drift and the diffusion may depend of the frequency vector x, allowing retro-actions of population on the environment itself.

The well posedness of the process (5.7), (5.8) is assured by classical arguments (see [49, 7]). Monte-Carlo simulations do not require substantial changes compared to the non-spatial case: the additional Brownian motion can be effectively simulated in a standard way.

5.3 A Fokker-Plank equation for the probability density

The stochastic process (5.7), (5.8) can be described in a deterministic way by means of two Kolmogorov integro-partial differential equations: the backward one, also known as Feynman-Kac equation (related to expected value), and the forward one, also known as Fokker-Plank equation (related to the density).

With minor changes from [4], one easily sees that the infinitesimal generator of the process (5.7) (settled in Σ^d), (5.8) is

$$\mathcal{L}\phi = a \cdot D_x\phi + v \cdot D_y\phi + \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^t D^2_{yy}\phi\right) + \mathcal{J}\phi. \tag{5.9}$$

Here D_x and D_y stand for the vectors of first derivatives w.r.t. $x \in \mathbb{R}^{d-1}$ and $y \in \mathbb{R}^N$, respectively, D^2_{yy} stands for the $N \times N$ matrix of the second order derivatives w.r.t. y, a, v, σ are the same functions appearing in (5.7), (5.8), and $\mathcal{J}$ is a non-local functional related to a discrete measure:

$$\begin{aligned}
\mathcal{J}(x, y, \phi) &= \int_{\mathbb{R}^{d-1}} \left(\phi(x + z, y, t) - \phi(x, y, t)\right) d\mu_{x,y}(z), \\
\mu_{x,y}(z) &= \sum_{\substack{i,j=1 \\ i\neq j}}^{d-1} \lambda_{ij} f_i(x, y)\delta_{\{\gamma_{ij}x_i(e_j - e_i)\}}(z) + \sum_{i=1}^{d-1} \lambda_{id} f_i(x, y)\delta_{\{-\gamma_{id}x_i e_i\}}(z) \\
&\quad + \sum_{i=1}^{d-1} \lambda_{di} f_d(x, y)\delta_{\{\gamma_{di}(1-\sum_{k=1}^{d-1} x_k)\}}(z).
\end{aligned}$$

The expected value at time t of a population which is at state (x, y) at time $t = 0$ is described by $u(x, y, t)$, the solution to the Feynman-Kac system

$$\begin{cases}
\partial_t u_k - a \cdot D_x u_k - v \cdot D_y u_k - \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^t D^2_{yy} u_k\right) = \mathcal{J}u_k, \\
u_k(x, y, 0) = \begin{cases} x_k & \text{as } k = 1, \dots d-1, \\ y_{k-d} & \text{as } k = d, \dots d+N-1. \end{cases}
\end{cases} \tag{5.10}$$

Otherwise, one can be interested into the macroscopic function $\varrho(x,y,t) \in [0,1]$, measuring the probability of finding a population distribution $(x_1, \dots x_{d-1}, 1-\sum_{i=1}^{d} x_i) \in \mathcal{S}^d$ in the position $y \in \mathbb{R}^N$ at time t. For instance at time $t > 0$ the quantity

$$P_i(t) = \iint\limits_{(B_\varepsilon(e_i)\cap\Sigma^d)\times\mathbb{R}^N} \varrho(x,y,t)dxdy$$

depicts the probability of having a high proportion of individuals of type i, while

$$P_i(t,\delta) = \iint\limits_{(B_\varepsilon(e_i)\cap\Sigma^d)\times B_\delta(0)} \varrho(x,y,t)dxdy$$

depicts the probability of finding a high proportion of individuals of type i near at the origin.

This can be done if the starting point is one population with character x in the position y (that is the initial datum is a Dirac mass centered at (x,y)), or if the initial status is a random variable with density function $\varrho_0(x,y)$. A rigorous deduction of the Fokker-Plank equation requests a-priori regularity of the density function. The topic of regularity can be addressed by the classical Hormander theory (see, for instance, the book [74]) and requests some technical assumptions, also in the diffusive setting (i.e. in absence of mutations). In the present setting there is no reason to expect that the density function is smooth enough, due to the anisotropy of diffusion and to the point process modeling mutation. We therefore choose to write the Fokker-Plank equation formally and then to settle it in the framework of viscosity solution theory. This approach has the advantage of asking very few a-priori regularity and producing well-posed solutions even in the degenerate elliptic, integro-differential setting arising from rare mutations.

Following Pavliotis [76] we compute $\mathcal{L}^*$, the dual operator in $L^2(\Sigma^d \times \mathbb{R}^N)$ of the infinitesimal generator:

$$\mathcal{L}^*\phi = \frac{1}{2}\sum_{h,k=1}^{N} \partial^2_{y_h y_k}\left((\sigma\sigma^t)_{hk}\phi\right) - \mathrm{div}_x\left(\phi a\right) - \mathrm{div}_y\left(\phi v\right) + \sum_{i=1}^{d} \mathcal{J}_i^*(f_i\phi), \tag{5.11}$$

where now

$$\begin{aligned}
\mathcal{J}_i^*(x,y,\phi) &= \int_{\mathbb{R}^{d-1}} \left(\phi(x+z,y,t) - \phi(x,y,t)\right) d\mu^i_{x,y}(z),\\
d\mu^i_{x,y}(z) &= \sum_{\substack{j=1\\ j\neq i}}^{d-1} \lambda_{ij}(1+\gamma^*_{ij})\, 1_{\Sigma^d}(x+\gamma^*_{ij}x_i(e_i-e_j))\delta_{\{\gamma^*_{ij}x_i(e_j-e_i)\}}(z)\\
&\quad + \lambda_{id}(1+\gamma^*_{id})\, 1_{\Sigma^d}(x+\gamma^*_{id}x_ie_i)\delta_{\{-\gamma^*_{id}x_ie_i\}}(z),
\end{aligned}$$

as $i = 1, \dots d-1$ and

$$\mu^d_{x,y}(z) = \sum_{j=1}^{d-1} \lambda_{dj}(1+\gamma^*_{dj})\, 1_{\Sigma^d}(x-\gamma^*_{dj}(1-\sum_{k=1}^{d-1}x_k)e_j)\delta_{\{\gamma^*_{dj}(1-\sum_{k=1}^{d-1}x_k)e_j\}}(z),$$

for $\gamma^*_{ij} = \gamma_{ij}/(1-\gamma_{ij})$.

It turns out that, if $\varrho_0(x,y)$ is the probability density of the random variable $X_0 = (x_0, y_0)$ describing the initial distribution of subpopulations, and if the solution $X_t = (x_t, y_t)$ to (5.7), (5.8) has a sufficiently smooth probability density $\varrho(x,y,t)$ for $t > 0$, then it solves the initial value problem

$$\begin{cases} \partial_t \varrho - \frac{1}{2} \sum\limits_{h,k=1}^{N} \partial^2_{y_h y_k} \left((\sigma\sigma^t)_{hk} \varrho \right) + \operatorname{div}_x (\varrho a) + \operatorname{div}_y (\varrho v) = \sum\limits_{i=1}^{d} \mathcal{J}_i^*(f_i \varrho) \\ \varrho(x,y,0) = \varrho_0(x,y), \end{cases} \tag{5.12}$$

in the closed set $(x,y) \in \Sigma^d \times \mathbb{R}^N$ and $t > 0$.

Let us explicitly remark that nonlocal operators $\mathcal{J}_i^*$ are not continuous w.r.t. x: this fact may have a huge instability effect. We therefore switch to another problem which is set into all $\mathbb{R}^{d-1} \times \mathbb{R}^N$ and is continuous. To this end we extend the fitness functions f_i, the drift v and the diffusion σ in a bounded smooth way to all $\mathbb{R}^d \times \mathbb{R}^N$ so that $f_i \geq 0$ have support contained in a cylinder, say $B_R(0) \times \mathbb{R}^N$. Concerning the initial datum ϱ_0, it can be extended as $\varrho_0 \equiv 0$ outside $\Sigma^d \times \mathbb{R}^N$. We thus look into the problem

$$\begin{cases} \partial_t \varrho - \frac{1}{2} \sum\limits_{h,k=1}^{N} \partial^2_{y_h y_k} \left((\sigma\sigma^t)_{hk} \varrho \right) + \operatorname{div}_x (\varrho a) + \operatorname{div}_y (\varrho v) + c\varrho = \tilde{\mathcal{J}} \varrho \\ \varrho(x,y,0) = \varrho_0(x,y), \end{cases} \tag{5.13}$$

for $(x,y) \in \mathbb{R}^d \times \mathbb{R}^N$ and $t > 0$, where now

$$\begin{aligned} \tilde{\mathcal{J}}(x,y,\phi) =& \int_{\mathbb{R}^{d-1}} (\phi(x+z,y,t) - \phi(x,y,t))\, d\mu_{x,y}(z), \\ d\mu_{x,y}(z) =& \sum_{\substack{i,j=1 \\ j \neq i}}^{d-1} m_{ij}(x,y) \delta_{\{\gamma^*_{ij} x_i (e_j - e_i)\}}(z) \\ &+ \sum_{i=1}^{d-1} m_{id}(x,y) \delta_{\{-\gamma^*_{id} x_i e_i\}}(z) + \sum_{j=1}^{d-1} m_{dj}(x,y) \delta_{\{\gamma^*_{dj}(1 - \sum_{k=1}^{d-1} x_k) e_j\}}(z), \\ m_{ij}(x,y) =& (1+\gamma^*_{ij}) \lambda_{ij} f_i(x + \gamma^*_{ij} x_i (e_i - e_j), y), \end{aligned}$$

as $i,j = 1, \dots d-1$, with $i \neq j$, and

$$\begin{aligned} m_{id}(x,y) =& \lambda_{id}(1+\gamma^*_{id}) f_i(x + \gamma^*_{id} x_i, y), \\ m_{di}(x,y) =& \lambda_{di}(1+\gamma^*_{di}) f_d(x - \gamma^*_{di}(1 - \sum_{k=1}^{d-1} x_k) e_i, y), \end{aligned}$$

as $i = 1, \dots d-1$,

$$c(x,y) = \sum_{\substack{i,j=1 \\ i \neq j}}^{d} \left(\lambda_{ij} f_i(x,y) - m_{ij}(x,y) \right).$$

It is worth clarify that the equation in (5.13) does not coincide with the one in (5.12) even if $x \in \Sigma^d$. Although they do coincide for that functions ϱ which are zero for x outside Σ^d. On the other hand if the support of ϱ_0 is contained in $\Sigma^d \times \mathbb{R}^N$ and $\varrho(t) \in L^1(\mathbb{R}^{d-1} \times \mathbb{R}^N)$ is nonnegative, then also the support of $\varrho(t)$ is contained in $\Sigma^d \times \mathbb{R}^N$.

To see this fact, let

$$\begin{aligned} A_k =& \{x \in \mathbb{R}^{d-1} : x_k < 0\} && \text{as } k = 1, \dots d-1, \\ A_d =& \{x \in \mathbb{R}^{d-1} : \sum_{k=1}^{d-1} x_k > 1\}, \\ I_k(t) =& \iint\limits_{A_k \times \mathbb{R}^N} \varrho(t) dxdy && \text{as } k = 1, \dots d. \end{aligned}$$

It suffices to check that $\dfrac{d}{dt} I_k(t) \le 0$. For simplicity we perform computations only in the case $d = 2$. Integrating the equation in (5.13) on $A_1 \times \mathbb{R}^N$ gives

$$\begin{aligned} \frac{d}{dt} I_1(t) = &- \int\limits_{\mathbb{R}^N} (a_1 \varrho)(0, y) dy + \lambda_{12} \iint\limits_{A_1 \times \mathbb{R}^N} \left((1 + \gamma_{12}^*)(f_1 \varrho)((1 + \gamma_{12}^*)x, y, t) - (f_1 \varrho)(x, y, t) \right) dxdy \\ &+ \lambda_{21} \iint\limits_{A_1 \times \mathbb{R}^N} \left((1 + \gamma_{21}^*)(f_2 \varrho)(x - \gamma_{21}^*(1 - x), y, t) - (f_2 \varrho)(x, y, t) \right) dxdy \end{aligned}$$

remembering that $a_1(0, y) \equiv 0$ and performing the obvious transformations in the second and third integrals yields

$$= -\lambda_{21} \int_{\mathbb{R}^N} dy \int_{-\gamma_{21}^*}^{0} dx (f_2 \varrho)(x, y, t) \le 0$$

because $f_2 \varrho \ge 0$. Similarly, since $a_1(1, y) \equiv 0$ one gets

$$\frac{d}{dt} I_2(t) = -\lambda_{12} \int_{\mathbb{R}} dy \int_{1}^{1 + \gamma_{12}^*} dx (f_1 \varrho)(x, y, t) \le 0.$$

It has also to be stressed that, in order to read the solution $\varrho(t)$ as a probability density, its total mass has to be 1, that is

$$M(t) = \iint\limits_{\mathbb{R}^{d-1} \times \mathbb{R}^N} \varrho(x, y, t) dxdy = 1 \quad \text{for all } t > 0,$$

provided that $M(0) = \iint\limits_{\Sigma^d \times \mathbb{R}^N} \varrho_0(x, y) dxdy = 1$. Again, integrating the equation in (5.13) gives

$$\begin{aligned} \frac{d}{dt} M(t) = &\lambda_{12} \iint\limits_{\mathbb{R} \times \mathbb{R}^N} \left((1 + \gamma_{12}^*)(f_1 \varrho)((1 + \gamma_{12}^*)x, y, t) - (f_1 \varrho)(x, y, t) \right) dxdy \\ &+ \lambda_{21} \iint\limits_{\mathbb{R} \times \mathbb{R}^N} \left((1 + \gamma_{21}^*)(f_2 \varrho)(x - \gamma_{21}^*(1 - x), y, t) - (f_2 \varrho)(x, y, t) \right) dxdy = \quad 0 \end{aligned}$$

after a trivial change of variables. Hence the total mass is preserved in the modified problem (5.13).

In view of these remarks, we can read as the probability density of the process (5.7), (5.8) a solution $\varrho(t)$ to the Cauchy problem (5.13) with the properties $\varrho(t) \in L^1(\mathbb{R}^{d-1} \times \mathbb{R}^N)$ and $\varrho(t) \geq 0$ for $t > 0$. The existence of such a solution is assured in the viscosity framework.

Theorem 21. *Assume that $f_i, v \in C^{1,1}(\mathbb{R}^{d-1} \times \mathbb{R}^N)$, $\sigma \in C^{2,1}(\mathbb{R}^{d-1} \times \mathbb{R}^N)$ are bounded together with their derivatives, with $f_i \geq 0$ and $\sigma \geq \varepsilon > 0$. Take ϱ_0 a Lipschitz-continuous, bounded function whose support is compact and contained in the interior of $\Sigma^d \times \mathbb{R}^N$ such that $\varrho_0 \geq 0$ and $\iint \varrho_0 dxdy = 1$. Then there exists a unique viscosity solution to (5.13). Moreover $\varrho(t) \in L^1(\mathbb{R}^{d-1} \times \mathbb{R}^N)$ and $\varrho(t) \geq 0$ for all $t > 0$.*

Proof. First of all the equation in (5.13) has to be written in the standard form of the viscosity solution framework, which is nonvariational. This can be done if the coefficients f_i, v, σ have the regularity requested by hypothesis. So we write

$$\partial_t \varrho + a\partial_x \varrho + b\partial_y \varrho + c\varrho - \frac{1}{2}\sigma^2 \partial^2_{yy} \varrho = \sum_{i=1}^{2} \tilde{\lambda}_i \mathcal{I}_i(\varrho) \tag{5.14}$$

where now

$$\begin{aligned} \tilde{\lambda}_1(x,y) =& \lambda_{12}(1+\gamma_1^*) f_1(x + \gamma_1^* x, y), \\ \tilde{\lambda}_2(x,y) =& \lambda_{21}(1+\gamma_2^*) f_2(x - \gamma_2^*(1-x), y), \end{aligned}$$

and consequently

$$c(x) = \partial_x a + \partial_y v - \frac{1}{2}\partial^2_{yy}\sigma^2 + \sum_{i=1}^{2} (\lambda_i f_i - \tilde{\lambda}_i)$$

are continuous and bounded. This problem satisfies the assumptions in [2], therefore it has a unique continuous viscosity solution $\varrho(x,y,t)$ which is Lipschitz-continuous w.r.t. x, y and bounded. Moreover, comparison principle holds. In particular one can find suitable parameters c_1, c_2, c_3 so that

$$0 \leq \varrho \leq \exp(c_1 t - c_2\sqrt{1+x^2} - c_3 y^2) \quad \text{in } \mathbb{R}^2 \times [0,\infty). \tag{5.15}$$

In particular $\varrho(t) \in L^1(\mathbb{R}^2)$ for all t. □

Remark 22. The assumption $\sigma \geq \varepsilon > 0$ has only been used to obtain the estimate from above in (5.15) and infer the integrability of the solution and the equation into all $\mathbb{R}^2$. The hypothesis can be removed by asking something more to the drift v in order to assure some decay w.r.t. y.

In view of the biological applications, it is suitable to allow the initial density ϱ_0 to be a probability measure. For instance modeling the evolution of one population whose initial state (x,y) is known deterministically requests to take a ϱ_0 as a Dirac mass centered at (x,y). This would hugely increase the mathematical difficulty. The paper [53] presents interesting results in this direction, which are modeled on the fractional Laplacian and therefore do not include the discrete non-local operator appearing here.

5.4 Hawks and Doves: a numerical study

In this section we take as a case study the two strategy game Hawks vs Doves ($d = 2$), with the following payoff matrix:

$$\mathcal{U} = \begin{pmatrix} \frac{G-C}{2} & G \\ 0 & \frac{G}{2} \end{pmatrix},$$

where the coefficients are both positive. The fitness functions for Hawks (x_1) and Doves (x_2), are respectively

$$f_1 = (G - C)x_1/2 + Gx_2, \quad f_2 = Gx_2/2,$$

then the replicator dynamics (reducing the coordinates only to $x \in [0, 1]$, fraction of Hawks) is

$$\dot{x} = x\,(1 - x)\,(f_1 - f_2) = x\,(1 - x)\,(G - Cx)\,/2.$$

Besides the pure-strategies equilibria $x = 0$ (all Doves) and $x = 1$ (all Hawks), a mixed strategies equilibrium can occur, $\bar{x} = G/C$, when $C > G$: in this case the real Hawks vs Doves game occurs, with $\bar{x}$ attractive and the other two values 0 and 1 which become unstable equilibria. Notice that when the cost of the fight C increases, the percentage of Hawks at the equilibrium $\bar{x}$ decreases. Instead, when the cost of fighting is less or equal than the gain, $C \leq G$, the only equilibria are the pure-strategies ones, with $x = 1$ attractive; the population tends to become only Hawks.

We add to the two strategies game also the space component, with $y \in \mathbb{R}$ ($N = 1$). In particular we assume that the cost for fighting depends on y as

$$C(y) = \frac{3G}{2}\left[1 + \frac{2}{\pi}\arctan(y)\right].$$

The function C is designed so that, at $y = 0$, the cost for fighting is $C = 3G/2 > G$ and we have a coexistence equilibrium $\bar{x} = 2/3$. At $y < 0$ the cost lowers untill it becomes equal to the gain for $y = -\sqrt{3}/3$, so for smaller values of y the coexistence equilibrium disappears, Hawks increase and the only attracting equilibrium is $\bar{x} = 1$. Otherwise if $y > 0$ environment is more favorable to Doves, because the cost increases up to $3G$, so that the fraction of Hawks at equilibrium $\bar{x}(y)$ is a decreasing function of y, tending towards $1/3$ as $y \to +\infty$. Summing up, for any fixed y, the standard replicator dynamics has its equilibrium at

$$\bar{x}(y) = \begin{cases} 1 & y < -\sqrt{3}/3, \\ G/C(y) & y \geq -\sqrt{3}/3, \end{cases} \tag{5.16}$$

which is well known to be a global attractor. In particular the initial state $(x_0, y_0) = (2/3, 0)$ is an equilibrium for the standard replicator dynamics (i.e. neither mutations or motions are allowed), and also when a deterministic motion with $v(2/3, 0) = 0$ is considered. All the simulations that follow represent the probability density $\varrho(x, y, t)$ evolving from the same initial state $(x_0, y_0) = (2/3, 0)$, and show that the equilibrium can be disrupted by Brownian motion in a heterogeneous environment and/or by mutations. They have been obtained in MATLAB using Monte-Carlo methods and, in the last section, using a numerical method for the I.P.D.E. (5.13).

5.4.1 Monte Carlo simulations.

Roughly speaking, large number of independent runs of the stochastic process is performed, to statistically estimate the density.

- Fixed the final time, T, we discretize the time interval $[0,T]$ in, at least, $N = 2^8$ sub-intervals with the same length. Fixed an accuracy α, the number N increases up to make sure that the probability of the event "up to one jump in each interval" is greater than $(1-\alpha)\%$;

- We choose the number of iterations of the method, itermax. We fix two values, N_x, N_y and the interval $[y_{min}, y_{max}]$ in which we want to display the density. Then we create a grid on $[0,1] \times [y_{min}, y_{max}]$, dividing the first interval in N_x parts, the second in N_y ($y_{min} = -5$, $y_{max} = 5$, $N_x = N_y = 50$). We define the array H in three dimensions, $N_x \times N_y \times N$, that will contain the following information:

$$H(i,j,t) = \frac{\#\left\{\text{processes s.t. at time } t \text{ are in the cell grid } (i-1,i)\times(j-1,j)\right\}}{\text{itermax}};$$

- For each iteration, we generate a Brownian motion on the N time points; then we generate a homogeneous Poisson process with intensity $\lambda_{max} \geq \max_x \lambda(x)$ on $[0,T]$. Let $\{T_1, \ldots, T_k\}$ be the jump times;

- We simulate, with Euler-Maruyama method, the stochastic process without jumps, until the nearest time T_i;

- Following the definition of the jump process and the intuitive interpretation presented before, we decide (acceptance-rejection) if the jump of the homogeneous process should be counted or not for the non-homogeneous one: if not, we continue Euler-Maruyama until the next jump. If so, we modify the population fractions in appropriate manner;

- We update the array H.

5.4.2 Numerical methods for the Fokker-Planck equation.

We implement a numerical method for the equations (5.13), that in this case has the form:

$$\begin{aligned}
&\partial_t \varrho - \frac{1}{2}\sigma^2 \varrho_{yy} + (\varrho a)_x + (\varrho v)_y = \mathcal{J}_1(\varrho,x,y) + \mathcal{J}_2(\varrho,x,y),\\
&\mathcal{J}_1(\varrho,x,y) = \lambda_{12}\left[\frac{1}{1-\gamma_{12}}(f_1\varrho)\left(\frac{x}{1-\gamma_{12}},y,t\right)1_{[0,1-\gamma_{12}]}(x) - (f_1\varrho)(x,y,t)\right],\\
&\mathcal{J}_2(\varrho,x,y) = \lambda_{21}\left[\frac{1}{1-\gamma_{21}}(f_2\varrho)\left(\frac{x-\gamma_{21}}{1-\gamma_{21}},y,t\right)1_{[\gamma_{21},1]}(x) - (f_2\varrho)(x,y,t)\right].
\end{aligned}$$

We obtain a finite differences scheme by discretizing with central difference the second order diffusive term and the transport term in y, and with a upwind method, that varies depending on the sign of the function a, for the transport term in x. The time is discretized using an explicit method. We denote with Δx and Δy the space steps, with Δt the time step, with

x_i, y_j the grid points and t_n the discrete times. Without considering the integral term (related to jumps) the method is conservative, and has the following form:

$$\begin{aligned}\frac{\varrho_{i,j}^{n+1}-\varrho_{i,j}^{n}}{\Delta t} &= \frac{\sigma^2}{2\left(\Delta y\right)^2}\left(\varrho_{i,j+1}^{n}-2\varrho_{i,j}^{n}+\varrho_{i,j-1}^{n}\right)\\ &\quad -\frac{1}{2\Delta y}\left((v\varrho)_{i,j+1}^{n}-(v\varrho)_{i,j-1}^{n}\right)\\ -\frac{1}{2\Delta x}\left[(a\varrho)_{i+1,j}^{n}-(a\varrho)_{i-1,j}^{n}-\left(\left|(a\varrho)_{i+1,j}^{n}\right|-2\left|(a\varrho)_{i,j}^{n}\right|+\left|(a\varrho)_{i-1,j}^{n}\right|\right)\right],&\end{aligned}$$

where $\varrho_{i,j}^{n}=\varrho(x_i,y_j,t_n)$.

Regarding the non-local jump terms $\mathcal{J}_1$, $\mathcal{J}_2$ the functions f_1, f_2 are well defined on non-grid points, but we have to approximate the value of ϱ^n in $x_i/(1-\gamma_{12})$, so we follow [3], using linear interpolation between the grid points $x_{\hat{i}}$ and $x_{\hat{i}+1}$, where $\hat{i}=\min\{j : x_j \le x_i/(1-\gamma)\}$:

$$\varrho^n\left(\frac{x_i}{1-\gamma_{12}}\right)=\frac{\left(\varrho_{\hat{i}+1}^{n}-\varrho_{\hat{i}}^{n}\right)}{\Delta x}\left(\frac{x_i}{1-\gamma_{12}}-x_{\hat{i}}\right)+\varrho_{\hat{i}}^{n}.$$

5.4.3 Replicator Dynamics perturbed by random motion.

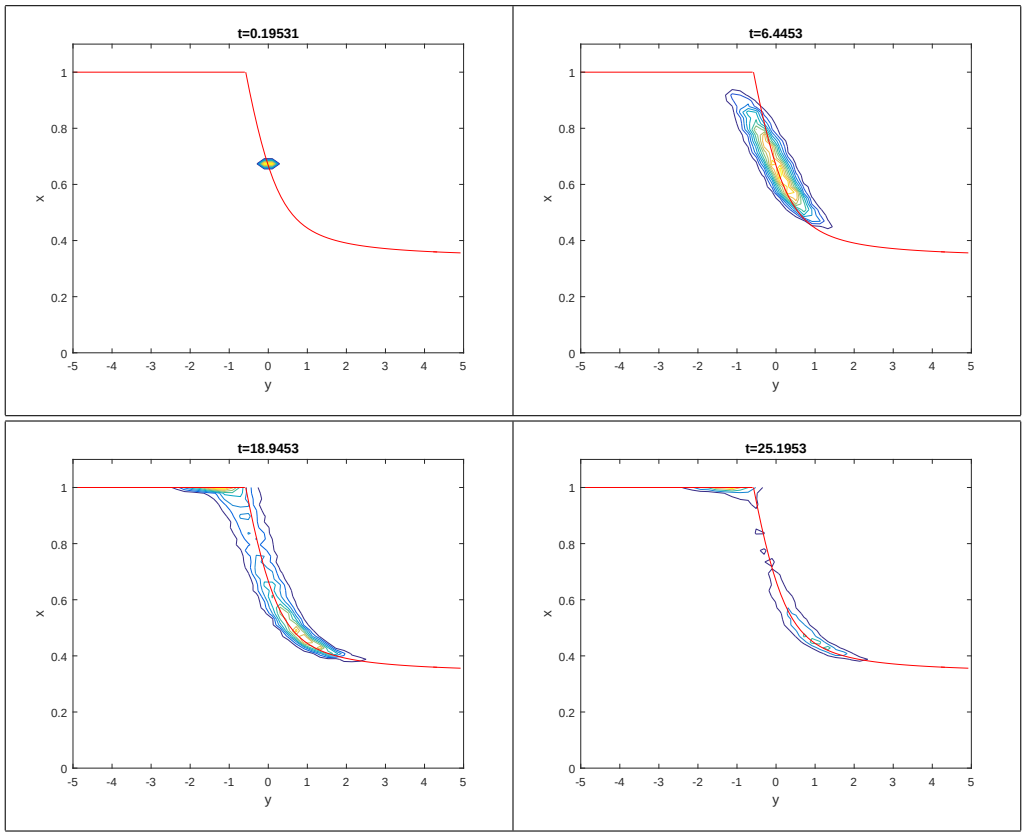


Figure 5.1: Replicator Dynamics perturbed by random motion, simulated by 10^5 iterations of the Monte Carlo method. The population moves randomly in space, subject to the selection of a changing environment. The red line is the function $\bar{x}(y)$, fraction of Hawks at the equilibrium for the standard Replicator Dynamics starting at y. The deterministic speed is zero, $v = 0$, jumps are absent, the coefficient of the Brownian motion is $\sigma = 0.2$. Other parameters: $T = 30$, $N = 2^8$, $y_{min} = -5$, $y_{max} = 5$, $N_x = N_y = 50$.

In this Monte Carlo simulation the population just moves randomly in space, subject to the selection of a changing environment. To do this, we imagine that jumps are absent, i.e. $K = 0$ in (5.7), and that (5.8) gives a homogeneous Brownian motion for the variable y, i.e. the drift v is zero and the diffusion coefficient is $\sigma = 0.2$. If the Brownian motion were absent, the character x_t of a population starting at (x_0, y_0) would tend as $t \to +\infty$ towards the attractor $\bar{x}(y_0)$ introduced in (5.16) and depicted by a red line in Figure 5.1. But now y_t follows (5.8), which reduces to a homogeneous Brownian motion, so that its marginal density is a Gaussian function with expected value y_0, kernel of the heat equation,

$$\varrho_{(y)}(y,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(y-y_0)^2}{2\sigma^2 t}\right\}.$$

Meanwhile the SDE (5.7) reduces to the standard replicator dynamics and moves x_t towards the asymptotically stable equilibrium $\bar{x}(y_t)$, which depends by y_t and therefore by time. We

can see how, with $t \gg 0$, the density is approximately

$$\varrho(x, y, t) \sim \bar{x}(y)\varrho_{(y)}(y, t)$$

with an expected global frequency of Hawks given by $\int_{\mathbb{R}} \bar{x}(y)\varrho_{(y)}(y, t)dy$, see Figure 5.1.

5.4.4 Replicator Dynamics plus Brownian motion with drift.

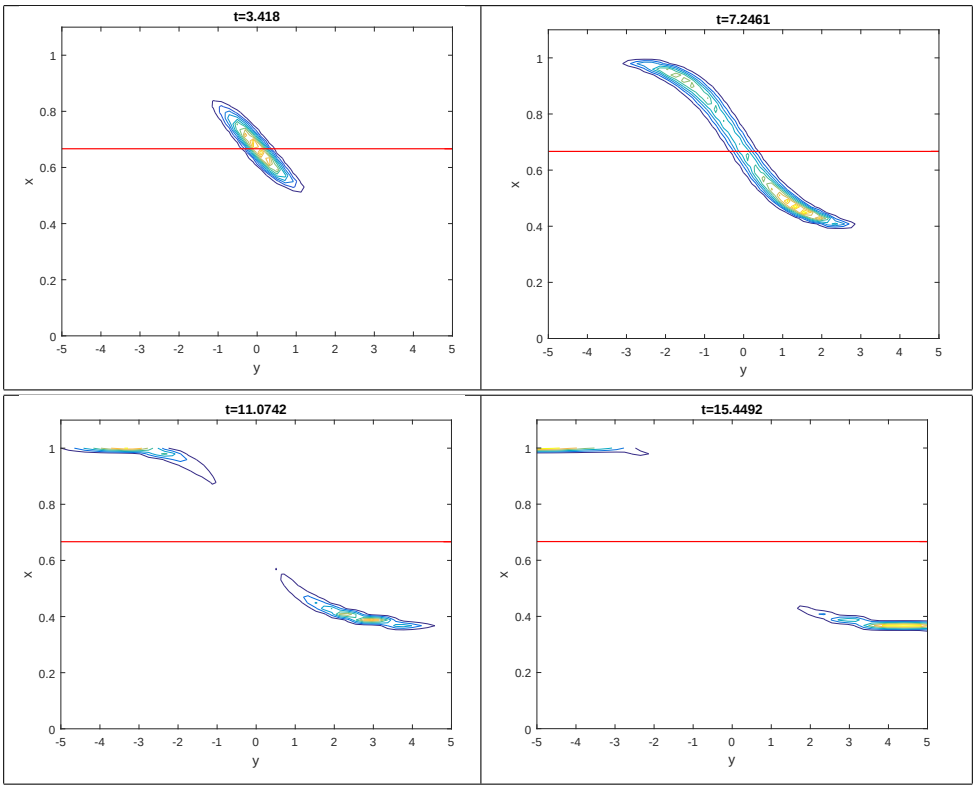


Figure 5.2: Replicator Dynamics plus Brownian motion with drift, simulated by 10^5 iterations of the Monte Carlo method. The vector x evolves according to the game Hawks vs Doves with G function of y. The red line represents the expected value of Hawks for $t \to \infty$, that is equal to the initial value of Hawks. The deterministic speed is chosen as $v = 1 - \frac{3}{2}x$, jumps are absent, the coefficient of the Brownian motion is $\sigma = 0.2$. Other parameters: $T = 30$, $N = 2^8$, $y_{min} = -5$, $y_{max} = 5$, $N_x = N_y = 50$.

We assume again that the character x_t follows the replicator dynamics with no jumps, i.e. we take $K = 0$ in (5.7). But now the we take a non-null drift in the environmental dynamics (5.8), depending on the character of the population:

$$v(x) = v_D\,(1 - x) + v_H x;$$

where $v_D > 0$ represents the drift of the Doves, moving towards positive values of y, and $v_H < 0$ the drift of the Hawks. When $v_D = 1$ and $v_H = -\frac{1}{2}$:

$$v(x) = 1 - 3x/2. \tag{5.17}$$

The drift is decreasing as a function of x (the proportion of Hawks): it has its maximum, $v = 1$, at $x = 0$ (high concentration of Doves) and its minimum, $v = -1/2$, in $x = 1$ (high concentration of Hawks). Moreover the drift is null at $x = 2/3$, which is taken as the initial state. As noticed at the beginning of this section, if the Brownian component were absent the initial state $(x_0, y_0) = (2/3, 0)$ would be an equilibrium and the resulting dynamics would be trivial. In the simulation depicted in Figure 5.2 the dynamics is not trivial, and we can identify two different behaviors of the process, because of the presence of the Brownian component with $\sigma = 0.1$. The support of the probability density function splits in two different regions, and it means that the population moves either towards negative values of y, or towards positive values, respectively with probability p_1 and p_2. In the first case, the proportion of Hawks at the equilibrium increases (as a function of $-y$), until the process oversteps the value $y = -\sqrt{3}/3$, after which $x = 1$ is the only equilibrium (all Hawks). We can see the gradual extinction of each Dove. In the other case, the cost of the fight increases with time, the density tends to concentrate toward the coexistence of both strategies, with greater concentration of Doves ($x = 1/3$). However, the expected value of the proportion of Hawks for $t \to +\infty$ (highlighted by a red line in Figure 5.2) is the same as the initial one, $2/3$. In fact, as we can see numerically, each of the two regions have mass $1/2$, so $p_1 = p_2$ and

$$\lim_{t \to +\infty} \mathbb{E}[x_t] = 1 \cdot p_1 + \frac{1}{3} \cdot p_2 = \frac{2}{3}.$$

5.4.5 Point-type mutations plus Brownian motion with drift.

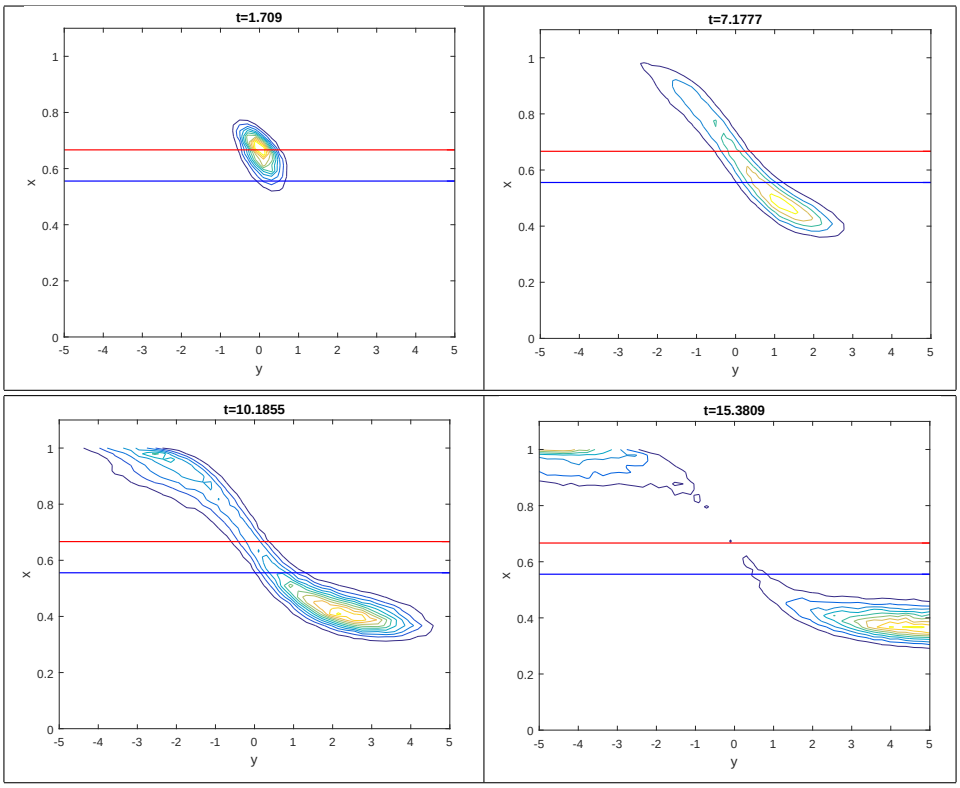


Figure 5.3: Point-type mutations plus deterministic and Brownian motion, simulated by 10^5 iterations. The vector x evolves according to the game Hawks vs Doves with G function of y. The red line, as in Figure 5.2, represents the initial value of Hawks and the expected value of Hawks for $t \to \infty$ in absence of mutations. The blue line is the same expected value for $t \to \infty$ in presence of mutations. The deterministic speed is chosen as $v = 1 - \frac{3}{2}x$, Brownian motion has $\sigma = 0.2$, the parameters of the jump process are $\lambda_{12} = \lambda_{21} = 0.2$, $\gamma_{12} = \gamma_{21} = 0.1$, that is one tenth of the population mutate each jump and we have "fair jumps". Other parameters: $T = 30$, $N = 2^8$, $\alpha = 0.1$, $y_{min} = -5$, $y_{max} = 5$, $N_x = N_y = 50$.

We take now a point-type mutation process for x_t, with $\lambda_{12} = \lambda_{21} = 0.2$; $\gamma_{12} = \gamma_{21} = 0.1$ in (5.6). Concerning motion, we take here $\sigma = 0.2$ and v given by (5.17), so that the position y_t changes deterministically with speed v and stochastically because of the Brownian motion. Let us remark that at each time that a mutation occurs, the probability that Hawks (respectively Doves) suffer a mutation only depends on fitness. At the initial state $(2/3, 0)$, the probability that Hawks are the first to suffer mutations is $1/2$, just like Doves. In this sense mutations produce random perturbations similar to the Brownian motion introduced in the previous example (5.4.4). The simulations presented in Figure 5.3 show two different regions also in this case. It is remarkable that the fact that at the equilibrium Hawks are more abundant than Doves brings as a consequence that mutations favor Doves, so that the region of the probability density moving rightwards will have higher mass (the ratio between right region and left region is 2:1 ca.), i.e. the coexistence of both strategies occurs with higher probability

($p_1 < p_2$), unlike example (5.4.4). The expected value of the proportion of Hawks for $t \to +\infty$ (highlightened by a blue line in Figure 5.3 is lower than the initial one:

$$\lim_{t \to +\infty} \mathbb{E}[x_t] = 1 \cdot p_1 + \frac{1}{3} \cdot p_2 \approx \frac{5}{9}.$$

We therefore see that including the physical space can favour the persistence of the low-fitness strategy, when mutations can happen in both directions.

5.4.6 Monte Carlo and Finite Differences simulations

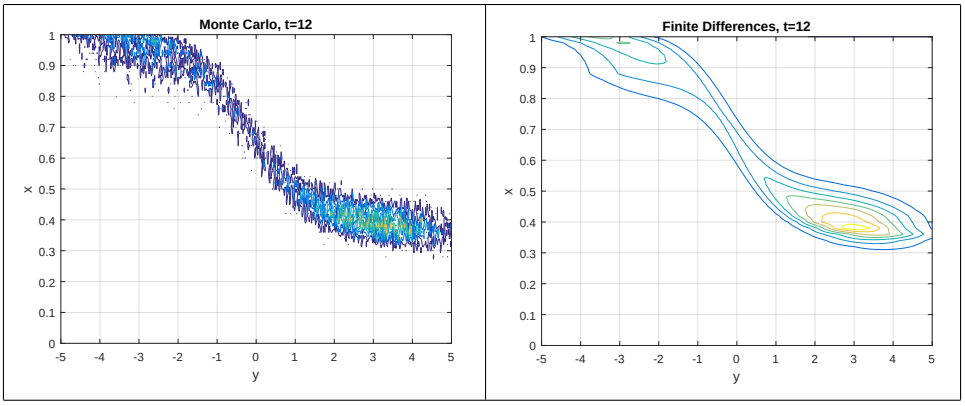


Figure 5.4: Replicator Dynamics plus Brownian motion, with drift and jumps, with different algorithms, both at time $t = 12$. On the left the graphic for Monte Carlo method, simulated by 10^4 iterations (the grid is dense, to show the single runs). On the right the scheme for the partial difference equation, with $\Delta x = \Delta y = 0.02$.

Here we compare the Monte Carlo algorithm, showed in the previous cases, and the finite differences approximation, presented at the beginning of Section (5.4). Let us take as study case the same problem as above, in Subsection (5.4.5). After the flattening of the initial datum (Dirac delta), which is slightly faster in the I.P.D.E. approximation, the two simulations run parallel (see Figure 5.4) creating the two regions with different masses, moving towards opposite directions. In the I.P.D.E. case, as in the Monte Carlo one, the ratio between the right region and left region masses is 2:1 ca.

Even if we choose a thick grid for the I.P.D.E. algorithm (500 cells of size Δx, 1000 cells of size Δy, 6000 time steps), its execution is about 5 times faster than the Monte Carlo simulation with 10^4 iterations. However, despite the speed, this algorithm, approximating the non-local jump term with linear interpolation, does not preserve the mass for the whole simulation. In a simulation, the mass at time $T = 15$ is 90% ca. of the initial one. Then, the choice of a singular initial point as a Dirac delta causes numerical dissipation in the transport terms, especially noticeable in simulations in which the Brownian motion is absent. For these reasons, Monte Carlo simulations have been privileged in Sections 5.4.3, 5.4.4 and 5.4.5.

Chapter 6

A Genotype-Phenotype model with mutations

In this chapter we build the equations for a new model of population dynamics, which follows over time a group of individuals undergoing Darwinian evolution. The basic models that inspired the new one are those of replicator dynamics with random mutations [4], in which a system of differential equations, deterministic and stochastic, is used to represent the different "types" in the population. The base of the new model is in the concept of "difference between genotype and phenotype". The genotype is the set of characteristics that define an individual, encoded in the DNA; a "type" of individuals is a group that shares the same genotype, and in the model there is an equation for each individual group in the population. Each genotype expresses a very specific phenotype, and the same phenotype can be the expression of several different genotypes, on which natural selection acts. The two main forces of Evolution, selection/adaptation and mutation, act at different levels: the environment select the fittest individuals, acting on the phenotype. The source of evolutionary novelty, or mutation, acts instead by randomly modifying the genotypes, producing new and never seen before ones.

In this chapter, after a section of motivations (6.1), we begin to construct a model, which is completed in Section 6.9. Growth equations, such as the Replicator dynamics (1.2), on which this model is based, are formulated for the frequencies of individuals in a population of multiple "types". These equations are reformulated in Section 6.2 to take into account the actual number of individuals. If the population is composed of an infinite number of types, then a new mathematical concept of population is necessary, since a symplex $\mathcal{S}^D$ takes into account only (at most) D different species, just as it is necessary to reformulate the previous equations in an adequate space. This is what is done in Section 6.3. In the following sections we discuss the concept of "type"; the previous equations were defined for generic "types". In this model the types, the unknowns of the equations, are the *genotypes*; the concept of potential space and population of genotypes is provided in Section 6.4. Equations can be written for genotypes only if fitness has a specific form. In Section 6.5 the concept of *fitness* is linked to that of *phenotype*, expression of a given genotype; the passage between genotype and phenotype is given by GP-map. If the *selection* is the force that acts on the phenotypes, through fitness, *mutations* totally concern the genotypes, and they act by ignoring the type of phenotype that a genome will express. After a section (6.7) dedicated to further theory of jump processes, the equations are then completed in Section 6.8, building a stochastic jump process. A proof of

existence and uniqueness of the process is given in Section 6.9.

See Figure 6.1 for a scheme of the Chapter.

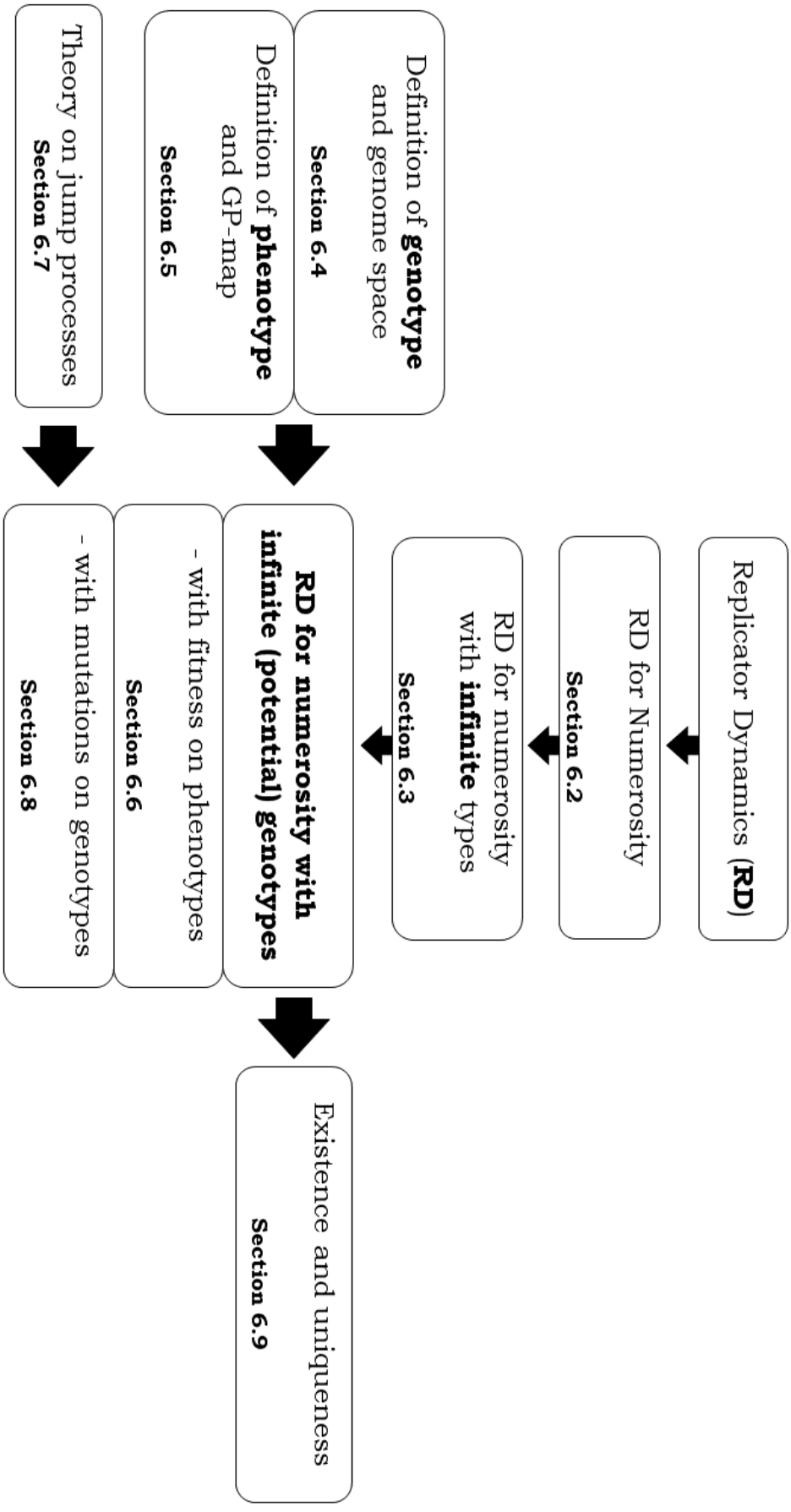


Figure 6.1: Diagram of the structure of Chapter 6 on the construction of the model.

6.1 Basics and motivations.

As we have seen in Part I, the basic model of replicator dynamics [73, 44] represents the trend of a population divided into D different types, $E_1, \ldots, E_D$, with relative frequency $x = (x_1, \ldots, x_D)$ in the simplex $\mathcal{S}^D$. The law governing their dynamics is

$$\begin{cases} \dot{x}_k = a_k(x) = x_k \left(\Phi_k(x) - \bar{\Phi}(x)\right), \\ k = 1, \ldots, D, \end{cases}$$

with Φ_k fitness function of the type k and $\bar{\Phi}$ average fitness in the population:

$$\bar{\Phi}(x) = \sum x_k \Phi_k(x).$$

with the structure of the fitness function that changes from context to context, and can be, according to the evolutionary game theory approach [44, 57], linked to a payoff matrix $\mathcal{U}$:

$$\begin{pmatrix} \Phi_1(x) \\ \vdots \\ \Phi_D(x) \end{pmatrix} = \mathcal{U}x.$$

The replicator dynamics model turns into a stochastic process when we consider random mutations that can transform a fraction of one type of individuals into another, through a series of Poisson processes (one for each pair of types (i, j) with $i \neq j$), non-homogeneous and with intensity proportional to a rate related to the fitness of the mutating type and to the "genetic distance" between mutant and mutated [4]. We refer from now to the model presented in [4] in the following way:

$$X_t = X_0 + \int_0^t a(X_s)\, ds + \sum_{i \neq j} \int_0^t \gamma(X_{s^-}) N_{ij}(ds). \tag{6.1}$$

Motivations

In the following we analyze the main restrictions of (6.1), that are the basis for the main novelties of the new model.

As M.Nowak writes in the introduction of his book [73], "the basic building blocks of evolutionary dynamics i.e. replication, selection and mutation apply to any biological organization anywhere in our or other universes and do not depend on the particular details of which chemistry was recruited to embody life. Any living organism has arisen and is continually modified by these three principles. More generally, these mechanisms are functional to any type of life, biological or not." From this perspective, the presented model is built not to be a phenomenological model, but a theoretical model of biological evolution, a "skeleton", which must be covered each time with the most suitable application. In the following, we will define the concepts of *space of the potential genomes*, of *"distance" between genomes* that is linked to mutations, of *genotype-phenotype map*, of *fitness function,* which regulates the interaction between different phenotypes.

Genotype and phenotype. The concept of "genotype-phenotype difference" is the base of the new model. As in [4], each equation represents a "type" of individuals, where in this case "type" means "genotype", then in the new model there is an equation for each group of individuals that shares the same genotype. A well-defined phenotype corresponds to each genotype, on which natural selection acts. The main forces of Evolution, selection, adaptation and mutation, act on different layers: the environment acts on the phenotype, selecting the fittest, while the randomness of the mutations affects the genotype. To show the differentiated effect, in the new model we have a system of equations for the genotypes, similar to the replicator dynamic, but with the fitness that is function only of the phenotypes. Mutations are random and act on genotypes, regardless of the phenotype expressed by that genotype. Two individuals with a "similar" phenotype have genotypes that can be very different. Similarly, a genotype that mutate is, with high probability, very similar to its predecessor, but the phenotypes may be very different.

We need to be clear about the definitions used from here for genotype and phenotype, as in [1]:

"Reproduction, as a fundamental property of biological systems, depends on the storage, processing and transfer of biological information. That information is typically stored in the form of sequences, such as DNA, RNA or amino acid sequences, and is more generally referred to as the genotype. In abstract models of biological evolution, the genotype can take more general forms, such as for genetic algorithms [65], where the genotypes are often binary strings. Genotypes are almost always linear and discrete representations of biological information." By GENOTYPE we mean in a general sense the *information* of an organism, which is replicable and modifiable due to the random processes of mutations. We do not require the level of detail that brings each individual to present a different genotype from each other, then we will group more individuals with similar genotype in the same quasispecies.

"Instead, the definition of a phenotype is almost impossibly wide"[1]. At any level of resolution of biological structure, any higher-level outcome resulting from a sequence can be seen as a phenotype: the amino acid to which a given triplet codon maps can be considered a phenotype. The structure and interactions of RNA and proteins are phenotypes, and, on a longer time scale, the development of an organism can also be seen as a phenotype, as well as its interaction with its ecosystem. Each organism actually has a phenotype a bit different from any other. However, in this model by PHENOTYPE we intend an organism as *a function of the selective pressures that act on it.* If two organisms have different features, but the selection forces do not notice these differences, then in the model those two organisms have the same phenotype.

To highlight the difference between genotype and phenotype, we use in this model the concept of Genotype-Phenotype map (GP map), a function that associates to each genotype the relative phenotype [1].

One of the most immediate connection between genotype and phenotype can be shown with the study of the protein folding problem [19], the spatial rearrangement of an amino acid sequence into a protein structure. It has been known since Mendel [63] that genetic mutations can cause phenotypic changes. In the early twentieth century, Fisher, Haldane and Wright produced the modern evolutionary synthesis [27, 39, 94], introducing the idea of fitness landscape [94], building on Mendel's and Darwin's [17] work. "A fitness landscape relates the space of genotypic variation to survival. It, therefore, contains a GP map implicitly. But the fitness landscape really consists of two distinct mappings: one is the GP map, and the other

is the mapping from phenotypes to fitness values.[1]"

Kimura [51] takes a step forward in this direction postulating that many mutations that are important for evolution must be neutral, meaning that they do not affect the fitness of the phenotype. Neutrality refers to selection rather than phenotypic change, and while it is possible for different phenotypes to be equally fit, it is reasonable to suppose that many mutations are not only neutral in terms of selection, but leave the phenotype entirely unchanged. This idea was substantially extended by Maynard Smith in 1970 [59], who addressed the apparent contradiction [81] between the vast number of possible amino acid sequences and the tiny fraction of these sequences that give rise to the proteins observed in nature. Maynard Smith postulates that "functional proteins must form a continuous network which can be traversed by unit mutational steps without passing through non-functional intermediates" [59], which is similar to the definition of what is now commonly referred to as a neutral network. Such networks mean that functional proteins occupy connected subsets of genotype space, which makes their discovery through an evolutionary process, driven by random mutations, feasible. "Maynard Smith arguably laid the foundations for the modern study of GP maps by proposing the concept of a protein space of all possible amino acid sequences, in which neighbours are defined by single amino acid substitutions."

GP maps have been analyzed from a statistical point of view, highlighting their structural characteristics and fundamental properties [36, 1]. In this model they are used in a dynamic way, to modify the replicator dynamics and to allow mutations and selection to act differently, see Section 6.4.

Infinite dimension of genome space and meaning of mutations. Another restriction in [4] concerns the number of potential genotypes that can be generated. Because the model lives on the simplex $\mathcal{S}^D$, it is not possible that more than D different types may exist in the population. This undermines the concept of mutation as a source of evolutionary novelty. In the new model we imagine that, from a population initially formed by a few types, with time and because of mutations, we can generate new genotypes, similar to those of the progenitors. A naive solution could be to define the process on a simplex of dimension D, very large, and define the initial data with only D_0 non-zero components, with $D_0 \ll D$, i.e. consider X_0 that lives on an appropriate simplex of dimension D_0. The replicator dynamics system evolves remaining on the simplex of size D_0, with the only changes that may lead to the formation of individuals with a "new" genome in the population.

Example 23. Let us take as an example the evolutionary game Rock-Paper-Scissors [44]. In this situation three types are in the population, R, P, S (Rock, Paper, Scissors), and the payoff matrix is

$$\mathcal{U} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix},$$

the average fitness is 0 while the equations for the fractions $x = (x_R, x_P, x_S)$ are

$$\begin{cases} \dot{x}_R = x_R\,(x_S - x_P)\,, \\ \dot{x}_P = x_P\,(x_R - x_S)\,, \\ \dot{x}_S = x_S\,(x_P - x_R)\,. \end{cases}$$

If we suppose x_0 with the third component null, that is $x_0 = \left(x_R^0, x_P^0, 0\right)$, i.e. the type "Scissors" does not exist in the population, and in the absence of mutations, the system quickly converges to the "Paper only" equilibrium (the system is reduced to the equivalent on $\mathcal{S}^2$);

$$\begin{cases} \dot{x}_R = -x_R x_P, \\ \dot{x}_P = x_R x_P. \end{cases}$$

If instead of the replicator dynamics we had taken into account the system (6.1), eventually the mutations would have generated the missing type, up to restore the balance between the three and give rise to the oscillating rock-paper-scissors dynamics. In this example, the "Scissors" type is a "novelty" in the population, generated as a mutation of already existing types.

This approach, however, shows a problem, namely the need to define *a priori* types not yet present in the population, almost completely nullifying the concept of novelty. Keeping in mind the example of DNA, it is possible that, starting from a fixed genome, the mutation can act by changing any base, or changing the length of the code, or in general by making changes that can potentially generate an infinite number of variants. For this reason, one of the fundamental steps is to define a space of potential, infinite and countable genomes that the system will explore step by step. This space contains all the possible genomes that can be created with the model, whether they are obtained or not. In the theoretical treatment of the process, the space of the genomes will therefore be known a priori, but in this case the concept of evolutionary novelty will become evident when the logic of numerical implementation of the model is shown. In fact, such logic is designed to mimic the steps of true biological behavior, and the desired effect of not having decided a priori where a mutation lead is achieved. A similar approach, without a genotype to phenotype mapping, can be found in [48], where a model of evolutionary dynamics is presented, in which every mutation leads to a different game characterized by a dynamic payoff matrix, growing and shrinking when a mutation occurs.

A limitation of models with mutations based on replicator dynamics is hidden in the form of the unknown variables: they represent fractions of individuals with mutations that are proportional to fitness, or to the fraction of individuals itself. By its nature, a mutation occurs randomly in the population, without depending on fitness or the fraction of individuals, generating a small group of newborns, small in the sense of absolute number. For this reason in the model we avoid to use fractions of individuals in the simplex $\mathcal{S}^D$ and begin to consider actual number of individuals N. In this case, the mutation of a type k does not change a fraction of individuals of that type but transforms a small number, possibly random, but independent of type k, unlike before, where it was an actual number dependent on the mutant, of individuals in new ones.

6.2 Number of individuals

As said in the previous section, models as the replicator dynamics, on which [4] is based, have the assumption that the population is constant in number. If this hypothesis is no longer valid, we must consider a new equation for the real number of individuals N_t, and transform the replicator equation (1.2), written for the frequencies, into an equivalent equation for the number. When, later in the text, the model is complete, the mutations will lead to the formation of small groups of individuals, of very small absolute number, regardless of how

large the total number of individuals in the population is. For this reason, the changes we make below are necessary. Hofbauer and Sigmund, in [44], show the following:

Lemma. *[44] There exists a differentiable, invertible map from $\mathcal{S}^D$ onto $\mathbb{R}_+^{D-1}$ mapping the orbits of the replicator equation*

$$\dot{x}^k = x^k\left(\Phi_k - \bar{\Phi}\right)$$

onto the orbits of the Lotka-Volterra equation,

$$\dot{n}^k = n^k\left(r_k + \sum_{i=1}^{D-1} a_{ki}n^i\right),$$

for $k = 1, \ldots, D-1$ and appropriate choices of r_k and a_{ki}.

Below, on the same steps, we present a way to transform the replicator dynamics on $\mathcal{S}^D$ in a model for the number on $\mathbb{R}_+^D$, adding a term of logistic growth in the population.

6.2.1 From number to frequencies

Let $N = \left(n^1, \ldots, n^D\right)$, $n^k \geq 0$ for $k = 1, \ldots, D$ be the vector of number of individuals in a population divided in D types (at this level we are not yet distinguishing between *geno*-types and *pheno*-types); with $X = \left(x^1, \ldots x^D\right)$, $x^k \geq 0$, for $k = 1, \ldots, D$, $\sum_k x^k = 1$ we indicate the fractions of individuals in the same population. We imagine that this population grows according to the following law:

$$\begin{cases} \dot{n}^k = F_k\left(N\right), \\ n^k(0) \geq 0, \\ k = 1, \ldots, D; \end{cases}$$

with $F = (F_1, \ldots, F_D)$ that assure existence and uniqueness of the differential equation. To get the equation for the frequencies, we change the coordinates as:

$$\begin{gathered} x^k := \tfrac{n^k}{\sum n^j}; \\ k = 1, \ldots, D \\ N^{tot} := \sum n^j. \end{gathered}$$

Carrying out the calculations:

$$\dot{N}^{tot} = \left(\sum \dot{n^j}\right) = \sum \dot{n}^j = \sum F_j(N) = \sum F_j(X, N^{tot});$$

$$\dot{x^k} = \left(\frac{\dot{n^k}}{\sum n^i}\right) = \left(\frac{\dot{n}^k N^{tot} - n^k \dot{N}^{tot}}{(N^{tot})^2}\right) = \frac{\dot{n}^k}{N^{tot}} - \frac{x^k \dot{N^{tot}}}{(N^{tot})^2} = \frac{F_k(X, N^{tot})}{N^{tot}} - \frac{x^k \sum F_i\left(X, N^{tot}\right)}{N^{tot}}$$

$$= \frac{1}{N^{tot}}\left(F_k - x^k \sum F_i\right) = x^k\left[\frac{1}{N^{tot}}\left(\frac{F_k}{x^k} - \sum F_i\right)\right].$$

We can define the fitness functions Φ_k:

$$\Phi_k = \Phi_k\left(X, N^{tot}\right) := \frac{F_k}{N^{tot}x^k}, \; \bar{\Phi} = \sum x^k \Phi_k = \sum \frac{1}{N^{tot}} F_k. \tag{6.2}$$

In conclusion we transform the initial equation into:

$$\begin{cases} \dot{x}^k = x^k \left(\Phi_k - \bar{\Phi}\right), \\ \dot{N}^{tot} = N^{tot}\bar{\Phi}, \\ x_0 = \left(x_0^1, \ldots, x_0^D\right) \in \mathcal{S}^D, \\ N_0^{tot} > 0; \end{cases}$$

Lotka-Volterra

The Lotka-Volterra growth equations (see Chapter 1), for the number of individuals in a population of prey (n^1) and predators (n^2),

$$\begin{cases} \dot{n^1} = \left(\alpha - \beta n^2\right) n^1, \\ \dot{n^2} = \left(-\gamma + \delta n^1\right) n^2, \end{cases} \tag{6.3}$$

are transformed into a replicator dynamics equation with fitness Φ_1 and Φ_2 that are functions of the number N^{tot} and the fractions x_1, x_2, according to (6.2):

$$\Phi_1\left(N^{tot}, x_1, x_2\right) = \alpha - \beta N^{tot} x_2,$$

$$\Phi_2\left(N^{tot}, x_1, x_2\right) = -\gamma + \delta N^{tot} x_1.$$

Defining $x_1 =: x \in [0,1]$, i.e. the fraction of prey, and consequently $x_2 = 1 - x$, the equation for fractions of prey (and number $N^{tot} \in \mathbb{R}^+$) is:

$$\begin{cases} \dot{x} = \left(\Phi_1 - \Phi_2\right) x(1-x) \\ \dot{N^{tot}} = N^{tot}\left[\Phi_1 x + \Phi_2 (1-x)\right] \end{cases} \tag{6.4}$$

The system for the number of individuals on $\mathbb{R}^+ \times \mathbb{R}^+$ admits the non trivial equilibrium:

$$(n_{eq}^1, n_{eq}^2) = \left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right),$$

that, for the system in $[0,1] \times \mathbb{R}^+$ becomes:

$$\left(N_{eq}^{tot}, x_{eq}\right) = \left(\frac{\alpha\delta + \gamma\beta}{\beta\delta}, \frac{\gamma\beta}{\alpha\delta + \gamma\beta}\right);$$

In Figure 6.2 it is possible to graphically visualize the trend of the two systems.

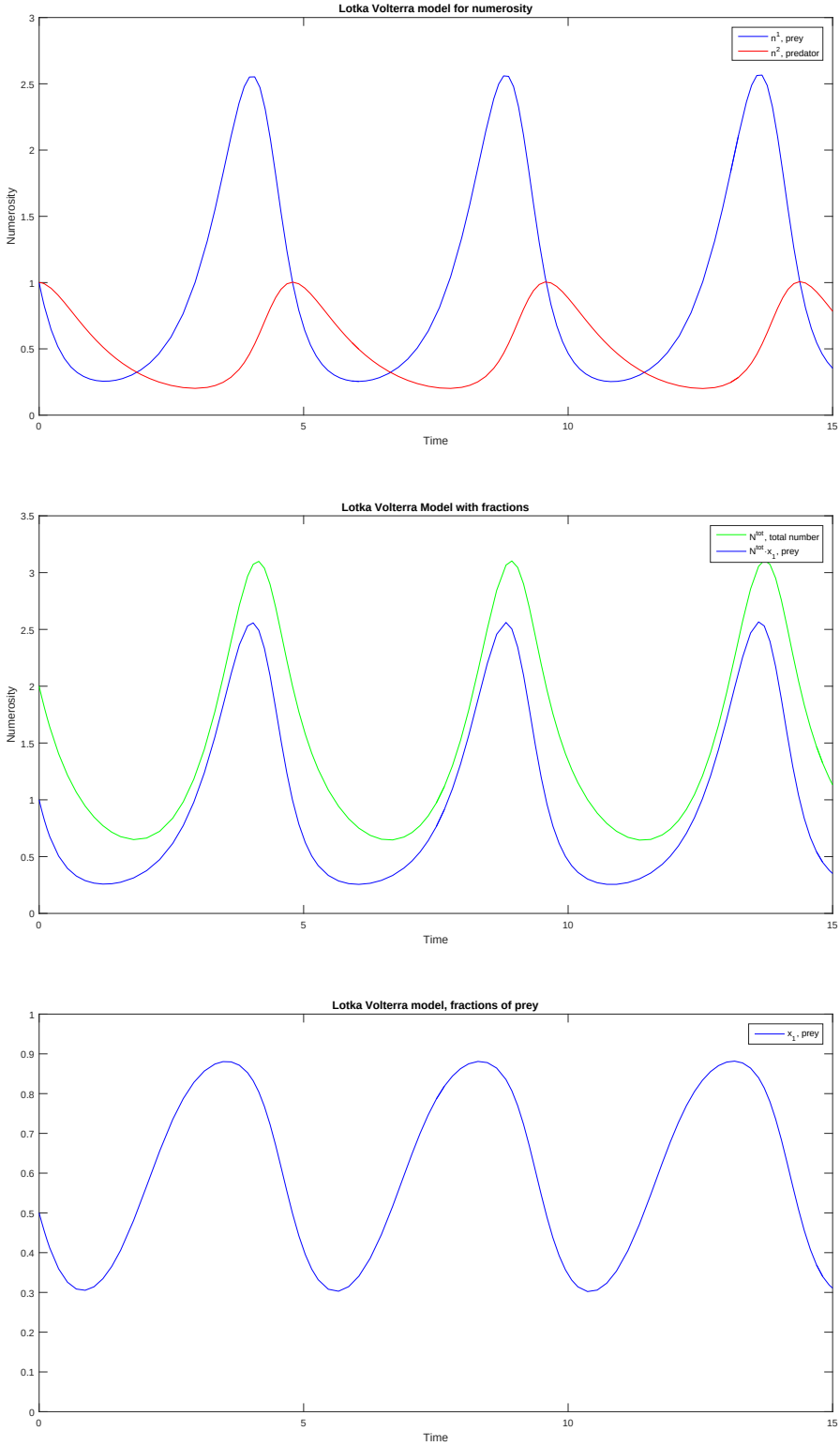


Figure 6.2: Lotka Volterra model with number of individuals and frequencies. In the first picture, the solution of the model for the number (6.3); in the second, the total number N^{tot}, and the number of prey $N^{tot}x_1$, computed from model (6.4); in the third the frequency of prey x_1. The initial datum is chosen $(1, 1)$ for model (6.3), i.e. $(2, 0.5)$ for model (6.4). Parameters are chosen as $\alpha = 2$, $\beta = 4$, $\gamma = \delta = 1$.

6.2.2 From frequencies to number

Let $X = \left(x^1, \dots, x^D\right) \in \mathcal{S}^D$ be the vector of frequency and $N = \left(n^1, \dots, n^D\right)$ the vector of number of individuals in a population divided in D types, as before. The system is modified according to the law of the replicator dynamics, plus an equation for the number N^{tot}:

$$\begin{cases} \dot{x}^k = x^k \left(\Phi_k - \sum_{i=1}^D x^i \Phi_i\right), \\ \dot{N}^{tot} = F\left(N^{tot}, X\right), \\ X_0 \in \mathcal{S}^D, \\ N_0^{tot} > 0. \end{cases}$$

To obtain the equations for the number, we change the coordinates:

$$\begin{gathered} n^k = N^{tot} x^k, \\ k = 1, \dots, D \\ N = \left(n^1, \dots, n^D\right). \end{gathered}$$

Carrying out the calculations:

$$\dot{n^k} = (\dot{N^{tot}} x^k) = \dot{N}^{tot} x^k + N^{tot} \dot{x}^k = F(N)\, x^k + N^{tot} x^k \left(\Phi_k(N) - \sum x^j \Phi_j(N)\right)$$

$$= F(N) x^k + n^k \left(\Phi_k(N) - \frac{1}{\sum n^j} \sum n^j \Phi_j(N)\right) = \frac{F(N) n^k}{\sum n^j} + n^k \left(\Phi_k(N) - \frac{1}{\sum n^j} \sum n^j \Phi_j(N)\right).$$

In conclusion we transform the initial equation into:

$$\begin{cases} \dot{n}^k = \frac{F(N)}{\sum n^j} n^k + n^k \left(\Phi_k(N) - \frac{1}{\sum n^j} \sum n^j \Phi_j(N)\right), \\ k = 1, \dots, D \end{cases} \tag{6.5}$$

6.3 Replicator dynamics on infinite space.

As we have seen in Section 6.1, we must set our problem in a framework with an infinite number of types, defining what is a *population* in this case, and how to generalize the equations (6.5). As we will see in the following Sections, in the model we can have a finite set of initial types that, with the passage of time, can generate new types, increasing the dimensionality of the problem. At each fixed time we will then have a finite set of types, but with cardinality that tends to (countable) infinity when time increases. We can then define a population:

Definition 24. In the space of all sequences of scalars $N = \left(n^1, \dots n^k, \dots\right)$, $n^k \in \mathbb{R}$ such that $n^k = 0$ except for finitely many n, let $\mathcal{S}_{\mathbb{R}}^{\infty}$ be the subset of the successions with compact support:

$$\mathcal{S}_{\mathbb{R}}^{\infty} = \left\{ N = \left(n^1, \dots, n^k, \dots\right) : \, n^k \geq 0 \, \forall k; \, \sum n^k < +\infty; \, D := \left| \left\{ k : \, n^k > 0 \right\} \right| < +\infty \right\}$$

A POPULATION is an element $N \in \mathcal{S}_{\mathbb{R}}^{\infty}$.

The population N is therefore a vector (infinite) that contains in position $k \in \mathbb{N}$ the number of individuals of type k. N is not really an infinite vector, because only a finite number of elements are non-zero (see Figure 6.3). If we imagine the number of individuals constant, without loss of generality equal to 1, $\sum n^k = 1$, then we can see how, in this case, $\mathcal{S}^\infty_{\mathbb{R}}$ is an "infinite union of symplexes", with a given element N that belong to a fixed symplex $\mathcal{S}^{D_N}$.

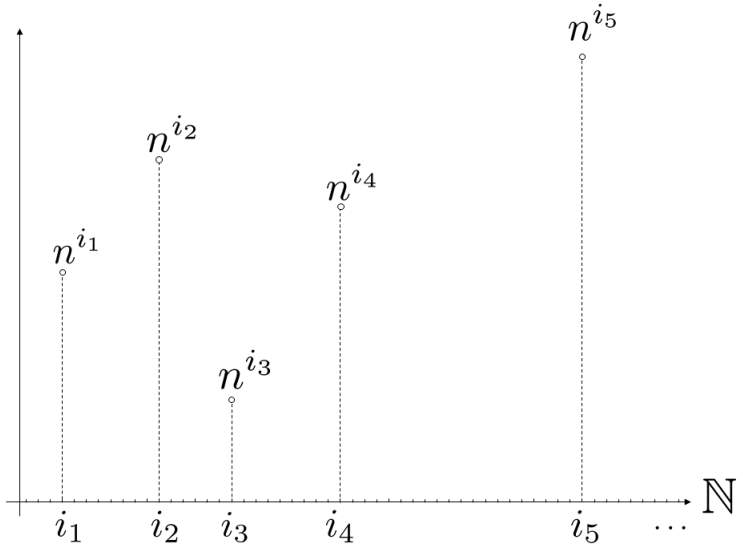


Figure 6.3: A population $N = (0, \ldots, n_{i_1}, \ldots, n_{i_2}, \ldots, n_{i_3}, \ldots, n_{i_4}, \ldots, n_{i_5}, \ldots, 0, \ldots) \in \mathcal{S}^\infty_{\mathbb{R}}$. At positions $i_k \in \mathbb{N}$ the numbers of individuals n^{i_k} are shown. In this case, only five of the n^k are not null.

On a population on $\mathcal{S}^\infty_{\mathbb{R}}$ we can define the model (6.5) and the fitness functions.
The fitness of the type k, for $k \in \mathbb{N}$ is a function Φ_k, with

$$\Phi_k : \mathcal{S}^\infty_{\mathbb{R}} \longmapsto \mathbb{R}.$$

The differential system (6.5) is then generalized, with $N_0 \in \mathcal{S}^\infty_{\mathbb{R}}$, as:

$$N_t = N_0 + \int_0^t a\left(N_s\right) ds, \tag{6.6}$$

$$a\left(N\right) = \begin{cases} \frac{F(N)}{\sum_{j\in\mathbb{N}} n^j} n^k + n^k \left(\Phi_k(N) - \frac{1}{\sum_{j\in\mathbb{N}} n^j} \sum_{j\in\mathbb{N}} n^j \Phi_j(N)\right), \\ k \in \mathbb{N} \end{cases}$$

Remark. It should be noted that in equation (6.6), given the choice of the initial datum, only a finite number of types is non-zero (only those in the initial datum support). The equations are therefore a system on an appropriate $\mathcal{S}^D$, masked. Switching to $\mathcal{S}^\infty_{\mathbb{R}}$ is justified when we want to add to the process terms that can change the number of types in the population, making them increase over time. This is done in Section 6.8; before that, in the next two sections the two concepts of Genotype and Phenotype are presented (and inserted in the logic of the model), which specifes the concept of "type" in the population, left intentionally generic up to this point.

6.4 Genotypes

Up to this point we indicated as *types* the groups of individuals on which the selection operates in the model (6.6), indicating a population of types with N (or with X in the case of fractions). As noted in Section 6.1, in this model we want to take a step forward and distinguish, at the level of mathematical construction, the role of genotype and phenotype.

The unknown variable of the complete model, that we continue to represent with N, indicates a "population of genomes". We imagine the genomes as elements of a large set that represents "all possible potentially existing genetic codes".

In Nature, a genome can, in the first instance, be represented by a sequence of finite length of nitrogenous bases, or a finite string of elements in the alphabet $\{A, C, T, G\}$. Mutations can alter the individual bases, leaving the length intact, or modify the code in a more substantial way, through deletions or insertions of bases. In this case, therefore, the space of the genomes will be represented by all the possible strings of finite length (not fixed) in the alphabet $\{A, C, T, G\}$. These requests lead us to the following definition:

Definition 25. POTENTIAL GENOME SPACE is a set $\mathcal{G}$, countable. An element $g \in \mathcal{G}$ is a GENOME

The concept of genome is the basis of the complete process, so we give justification of the choices in the definition. Since the process must represent mutations as a source of evolutionary novelty, the space of the genomes is explored starting from an initial genome through mutations. For this reason, the space $\mathcal{G}$ must represent the complete set of all possible variants of a given genome. In the realistic example of a genome as a string of nitrogenous bases, the possibilities are unlimited; from this consideration the choice of $\mathcal{G}$ not necessarily finite. On the other hand, in the model we want, to trace the type of mutations in DNA, that the variations are quantized, and there is not a continuous infinity of genomes; hence the choice of $\mathcal{G}$ at most countable.

Because the set $\mathcal{G}$ is countable, it is possible to define a sorting function α, so that it is possible to refer to the genomes with a number in $\mathbb{N}$. This, which seems a useful technicality for the writing of the model, has an important role in the logic of simulation, as we will see later.

Definition 26. We define the function $\alpha : \mathbb{N} \longmapsto \mathcal{G}$, which is an order on $\mathcal{G}$, so that it is possible to associate each genomes to a progressive number (i.e. mutation from the genome i to the genome j).

Using the definition of α is then possible to define a *population of genomes* as a point $N \in \mathcal{S}_{\mathbb{R}}^{\infty}$, indicating with n^k the number of individuals that shows the genotype $g_k := \alpha(k)$ (see Figure 6.4). Before defining the equations for the evolution of the population of genomes N, that are a variant of (6.6), it is necessary to identify what is the role of the phenotype in the model, and how, through a GP-map, genotype and phenotype are linked.

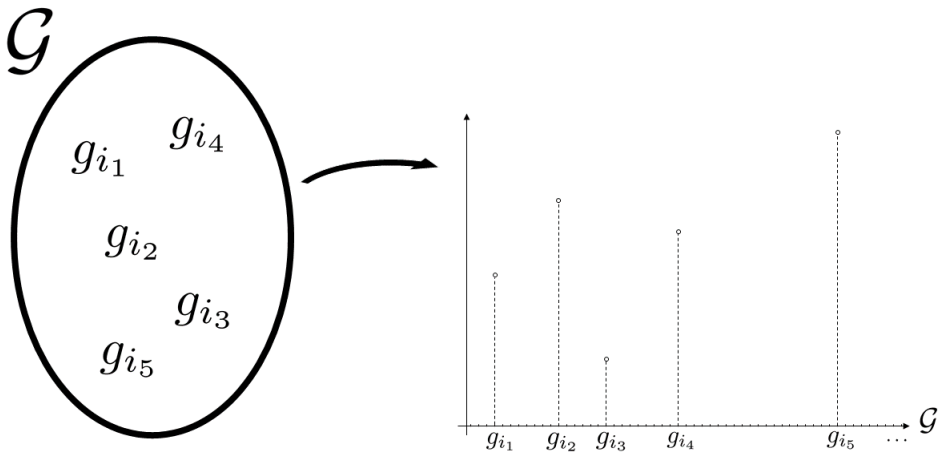


Figure 6.4: Starting from a genome space $\mathcal{G}$, sorting all the individuals, a population of genomes is defined, as in Figure 6.3. On the x axis, the genomes are shown, with their number on the y axis. Only a finite number of genomes have number not null.

6.5 Phenotypes and GP-maps

In our model, a phenotype must be associated with each genotype, expression of that given genotype. The genetic information that is stored in DNA, represented in its primary structure as a string of nitrogenous bases and in the set $\mathcal{G}$, is clearly not a random set of characters, just as a random set of genes does not generate a living organism. If a nucleic acid is mathematically represented as a string with letters in the alphabet A, C, T, G, strings which do not represent anything that is biologically valid can also be formed with the same letters. In addition, if we assume that mutations are a selection-blind force that acts by modifying the structure of DNA strings, the probability that, after a number (even high) of mutations, we can pass from a biologically valid string to a nonsense one it is greater than zero. The role of GP-maps is therefore transforming valid strings into individuals who will participate in the struggle for adaptation posed by the environment, and not to recognize as individuals the genomes that instead represent nothing (those individuals will be called non-viable);

We define then the phenotype space and genotype-phenotype map as:

Definition 27. POTENTIAL PHENOTYPE SPACE is a set $\mathcal{F}$, countable; an element $f \in \mathcal{F}$ is a *phenotype*. In $\mathcal{F}$ there may exist an element, which we indicate with $'0'$, called non-vital phenotypes. Given a genotype space $\mathcal{G}$ and a phenotype space $\mathcal{F}$ a GENOTYPE-PHENOTYPE MAP GP is a function:

$$GP : \mathcal{G} \to \mathcal{F}.$$

A GP map is not injective, because different genotypes may generate the same phenotype. The set of genotypes that produce a given phenotype is called the *Neutral Space* of that phenotype; in particular, $\mathcal{G}_0$ indicates the neutral space of $'0' \in \mathcal{F}$, i.e. $\mathcal{G}_0 = GP^{-1}('0')$.

The space of the phenotypes $\mathcal{F}$ is different from that of genotypes $\mathcal{G}$, and the Darwinian selection acts on the phenotypes. In our model, selection should not be able to distinguish two individuals that have the same phenotype, even if they have different genotypes. A phenotype is associated to a genotype through a GP-map and a fitness function is associated to a phenotype, as a function of the other genotypes in the population (see Figure 6.5).

$$\begin{array}{ccc} \mathbb{GENOTYPES} & \longmapsto & \mathbb{PHENOTYPES} \\ & GP-map & \uparrow \\ & & Fitness \end{array}$$

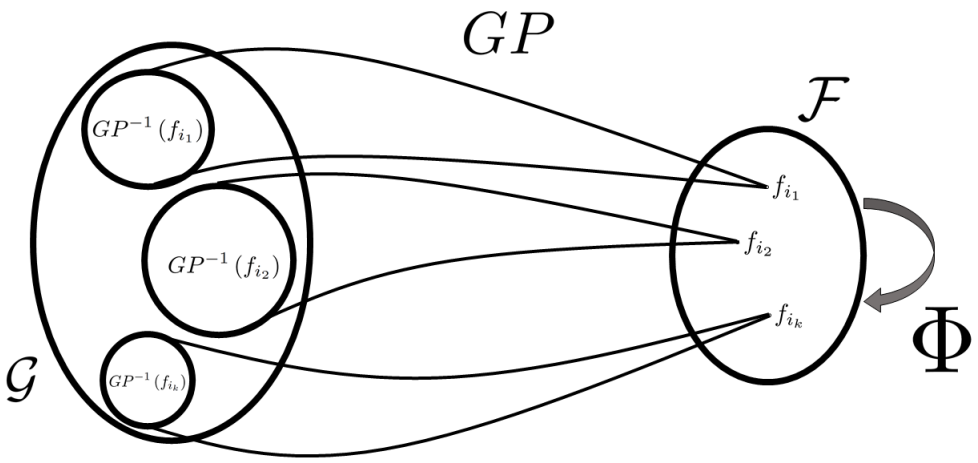


Figure 6.5: The phenotype-genotype map GP allows genotypes in $\mathcal{G}$ to express phenotypes in $\mathcal{F}$, and it is non-injective. The fitness function Φ acts exclusively on the phenotype space, ignoring $\mathcal{G}$. When added, in Section 6.8, mutations will act exclusively on the genome space $\mathcal{G}$.

6.6 Fitness for phenotypes

We propose here to insert the concepts of genotype and phenotype presented Sections 6.4 and 6.5 in the model on $\mathcal{S}_{\mathbb{R}}^{\infty}$ (6.6). Let us consider the genome space $\mathcal{G}$ and the population of genomes N on $\mathcal{S}_{\mathbb{R}}^{\infty}$. The GP-map GP defines a population of phenotypes $M \in \mathcal{S}_{\mathbb{R}}^{\infty}$:

$$M = \left(m^1, \ldots, m^k, \ldots\right) \in \mathcal{S}_{\mathbb{R}}^{\infty}$$

where m^k is the number of individuals of phenotype f_k.

$$m^k = \sum_{g_i \in \mathcal{G}:\, GP(g_i) = f_k} n^i;$$

on $\mathcal{F}-\{'0'\}$ we define the fitness function $\tilde{\Phi}=\left(\tilde{\Phi}_k\right)_{f_k\in\mathcal{F}-\{'0'\}}$, such that for each phenotype f_k:

$$\Phi_k : \mathcal{S}_{\mathbb{R}}^{\infty} \longmapsto \mathbb{R},$$
$$M \longmapsto \tilde{\Phi}_k(M).$$

The functions Φ_k depends on the other phenotypes and their abundance in the population, i.e. Φ_k is a function of M. Since we want the equations of the system to be for the population of genotypes N, we write the function $\tilde{\Phi}$ in the variable N obtaining a fitness function defined for each genotype in $\mathcal{G}$, except $\mathcal{G}_0$, function of the whole population of genotypes, i.e. $\tilde{\Phi}_k$ is a function of N. $\Phi=(\Phi_k)_{g_k\in\mathcal{G}-\mathcal{G}_0}$:

$$\Phi_k : \mathcal{S}_{\mathbb{R}}^{\infty} \longmapsto \mathcal{S}_{\mathbb{R}}^{\infty} \longmapsto \mathbb{R},$$
$$N \longmapsto M \longmapsto \tilde{\Phi}_k(M) = \Phi_k(N).$$

With this choice of fitness function, it is therefore possible to formulate the model (6.6) for each $g\in\mathcal{G}-\mathcal{G}_0$. As we have seen, only genotypes that express a vital phenotype ($g\in\mathcal{G}-\mathcal{G}_0$) interact according to the mechanism of Replicator Dynamics on $\mathcal{S}_{\mathbb{R}}^{\infty}$ (6.6), while we want all $g\in\mathcal{G}_0$ to rapidly extinguish, without interacting with the others. This group, once generated, must disappear in very short time. Then we imagine that these individuals decrease exponentially, so we add a decrease term for each of these,

$$\dot{n}^k=-\lambda n^k,\ \lambda\gg 0,\ k:\ g_k\in\mathcal{G}_0.$$

Overall, the differential system (6.6) is generalized for all $g\in\mathcal{G}$, with $N_0\in\mathcal{S}_{\mathbb{R}}^{\infty}$, as:

$$N_t=N_0+\int_0^t \tilde{a}(N_s)\,ds, \tag{6.7}$$

$$\tilde{a}(N)=\begin{cases}\frac{F(N)}{\sum_{j:\,g_j\in\mathcal{G}-\mathcal{G}_0} n^j}n^k+n^k\left(\Phi_k(N)-\frac{1}{\sum_{j:\,g_j\in\mathcal{G}-\mathcal{G}_0} n^j}\sum_{j:\,g_j\in\mathcal{G}-\mathcal{G}_0} n^j\Phi_j(N)\right), & k:\ g_k\in\mathcal{G}-\mathcal{G}_0\\ -\lambda n^k,\ \lambda\gg 0. & k:\ g_k\in\mathcal{G}_0\\ k\in\mathbb{N}\end{cases}$$

Remark. The model without mutations is unnecessarily complex. Until now, we have not been interested in mutations, and we have developed a model where selection and adaptation act on phenotypes. However, in the absence of mutations, the presence of infinite space is irrelevant. In fact, mutation is the force that generates new genotypes, consequently increasing the dimensionality of the problem. With no mutations, even if the population composition varies, the number of individuals does not change. Infact, let us consider the model (6.7) on $\mathcal{G}$. Since N_0 is in $\mathcal{S}_{\mathbb{R}}^{\infty}$, only a finite number of n_0^k are null, without loss of generality, $n^k=0$ for all $k>D$, with $\{g_0\}=\mathcal{G}_0$. There are no functions that increase the dimensionality of the problem, that is indistinguishable from a system on $\mathbb{R}^{D+1}$:

$$\dot{n^k}=\begin{cases}\frac{F(N)}{\sum_{j=1}^{D} n^j}n^k+n^k\left(\Phi_k(N)-\frac{1}{\sum_{j=1}^{D} n^j}\sum_{j=1}^{D} n^j\Phi_j(N)\right), & k=1,\dots,D\\ -\lambda n^0,\ \lambda\gg 0.\end{cases}$$

On the other hand, the absence of mutations makes the presence of the GP-map irrelevant too, since all the equations of genotypes in the same neutral space have the same form, and the model, from the system on $\mathbb{R}^{D+1}$ can be traced back to a system with the number of phenotypes as unknowns, on $\mathbb{R}^{d+1}$, with $d < D$:

$$\dot{m^k} = \begin{cases} \frac{F(M)}{\sum_{j=1}^{d} m^j} m^k + m^k \left(\Phi_k(M) - \frac{1}{\sum_{j=1}^{d} m^j} \sum_{j=1}^{D} m^j \Phi_j(M) \right), \quad k = 1, \dots, d \\ -\lambda m^0, \ \lambda \gg 0. \end{cases}$$

with

$$m^k = \sum_{g_i \in \mathcal{G}: GP(g_i) = f_k} n^i.$$

As stated in this remark, therefore, the construction made up to now is unnecessarily complex, unless there are mutations. In the next two sections we then construct the process of mutations, so we can justify the presence of infinite space, and we show that mutations act on the genotype, regardless of the type of phenotype that genotype expresses.

6.7 Jump processes, general interpretation

In Section 6.8 we construct the mutation process for the genomes. For this reason we premise this theoretical section where we develop the theory of jump process (refer to Chapter 2) reporting the details of the construction of a random Poisson measure and of some types of processes, from an "algorithmic" point of view. Moreover, reasoning in this optic will also be useful the understand more clearly the interpretation. In this section, we're going to consider processes of pure jump in this form:

$$X_t - X_0 = \int_0^t \int_E K(X_{s^-}, \xi) \mathcal{N}(ds, d\xi),$$

in $\mathbb{R}^D$, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a right-continuous filtration $\{\mathcal{F}_t, t \geq 0\}$. K is Borel measurable function, $\mathcal{N}(ds, d\xi)$ is a Poisson random measure on $\mathbb{R}^+ \times E$ with mean measure $l \times \nu$, l Lebesgue measure on $\mathbb{R}^+$, ν a σ-finite measure on a measurable space $(E, \mathcal{E})$. A jump process is then identified by the space of markers E and from the intensity function K. We are therefore going to build in detail both E and K, at a technical and interpretative level.

Remark. Since this section is purely theoretical, solutions X_t do not have any kind of biological meaning. They do not indicate, for example, fractions of individuals.

6.7.1 Poisson random measure

In Chapter 2 we introduced the concept of random Poisson measure and presented a first construction of it, taken from the book [50], when the measure of the marks space E was finite. Below we expand that construction, considering a generic random measure. After that we present a series of examples, which will lead to the construction of the final model.

Lemma 28. *[50] Let $\mathcal{N}(ds, d\xi)$ be a Poisson random measure on $\mathbb{R}^+ \times E$, with (E, ν) a measurable space, ν σ-finite measure.*

- *Suppose $\nu(E) < +\infty$. Define $\lambda := \nu(E)$ and let $(T_n)_{n\in\mathbb{N}}$ be the jump times of a homogeneous Poisson process on $\mathbb{R}^+$, with intensity λ. Let $(\xi_n)_{n\in\mathbb{N}}$ be independent random variables, identically distributed, uniform on (E, ν). Then, the random measure $\mathcal{N}(ds, d\xi)$ can be expressed as*

$$\mathcal{N}(ds, d\xi) = \sum_{n\in\mathbb{N}} \delta_{T_n}(ds)\delta_{\xi_n}(d\xi)$$

- *Suppose $\nu(E) = +\infty$ and use the σ-finiteness, so you can have a sequence (E_k) of disjoint sets, with $\bigcup E_k = E$ and $\nu(E_k) < +\infty$ for all k. On these sets, define $\{\nu_k = \nu\mathbb{I}_{E_k}\}$, $\nu_k(A) = \nu(A \cap E_k)$ and repeat the above process. Generate the sequence of random measure $\{\mathcal{N}_k\}$. The measure $\mathcal{N}$ can be obtained as sum of all the $\{\mathcal{N}_k\}$,*

$$\mathcal{N}(ds, d\xi) = \sum_k \mathcal{N}_k(ds, d\xi).$$

When the measure $\nu(E)$ is finite, we can think a Poisson random measure as a succession of random points (T_n, ξ_n) in the space $\mathbb{R}^+ \times E$. The values in $\mathbb{R}^+$ are occurrences of a homogeneous Poisson that has intensity $\nu(E)$. For each event T_n there is a value ξ_n, uniformly sampled in E, with measure ν. Thanks to this, the role of the function K becomes clear, in fact, we can write:

$$\int_0^t \int_E K(X_{s^-}, \xi)\mathcal{N}(ds, d\xi) = \sum_{n:\, T_n \le t} K(X_{T_n^-}, \xi_n);$$

A jump process, therefore, at a fixed time $t \in \mathbb{R}^+$ calculates K on each pair (T_n, ξ_n) with $T_n \le t$ (see Figure 6.6 for a visual representation of the costruction in Lemma 28). To construct a process it is therefore necessary to establish precisely what the form of the function K is and what the space E is, a space of "marks", which can be shaped to give the desired structure to the process.

When the set of marks E has infinite measure, $\nu(E) = \infty$, the construction presented above can be repeated, provided that it has to be replicated for each subset of finite measure $E_1, \ldots, E_i, \ldots$, that form a partition of E: there will be a succession of pairs (T_n^i, ξ_n^i) for each subset E_i.

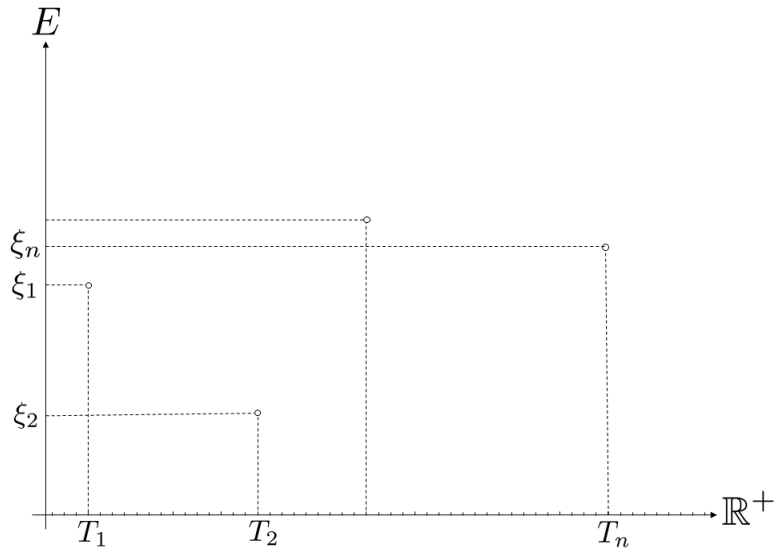


Figure 6.6: Visual representation of the costruction in Lemma 28 of a Poisson random measure $\mathcal{N}$ on $\mathbb{R}^+ \times E$ (with $\nu(E) < \infty$). On the x axis the jump times T_n, $n \in \mathbb{N}$ of a homogeneous Poisson process are reported; on the y axis the succession of marks ξ_n, $n \in \mathbb{N}$ in E. The function K acts on the space $\mathbb{R}^+ \times E$, i.e. on the pairs (T_n, ξ_n), $n \in \mathbb{N}$.

In the following, some example of simple processes:

Homogeneous Poisson process A homogeneous Poisson process of intensity λ is obtained by choosing as space $E = (0, \lambda)$, $\nu = Leb_{\mathbb{R}}$, e $K(x, \xi) = 1$. Times T_n are from a Poisson process with intensity $\nu(E) = \lambda$ and the marks ξ_n (uniform variables on $(0, \lambda)$) are unused in this case:

$$X_t - X_0 = \int_0^t \int_E K(x, \xi) \mathcal{N}(ds, d\xi) = \sum_{T_n < t} 1 = \#\{T_n < t\}$$

With the time t fixed, we count the pairs (T_n, ξ_n) with $T_n \leq t$.

Non-homogeneous Poisson process We choose to write now a jump process X_t with non-homogeneous intensity, function of the process itself, $\lambda = \lambda(X_{t^-})$. The function $\lambda(x)$ is positive and limited ($0 < \lambda(x) \leq \lambda_{max} < \infty$). This process is obtained by choosing as a state space $E = (0, \lambda_{max})$, as measure $\nu = Leb_{\mathbb{R}}$ and $K(x, \xi) = 1_{\{\xi < \lambda(x)\}}(\xi)$. With a process of acceptance-rejection, we consider a homogeneous Poisson process with higher frequency of jumps and, at each occurrence of a jump, we evaluate the function $\lambda(\cdot)$ before the jump. Only if the mark ξ is less than $\lambda(x)$ the jump of the homogeneous process is accepted and actually takes place (see Figure 6.7).

$$X_t - X_0 = \int_0^t \int_0^{\lambda_{max}} 1_{\{\xi < \lambda(X_{s^-})\}}(\xi) \mathcal{N}(ds, d\xi) \tag{6.8}$$

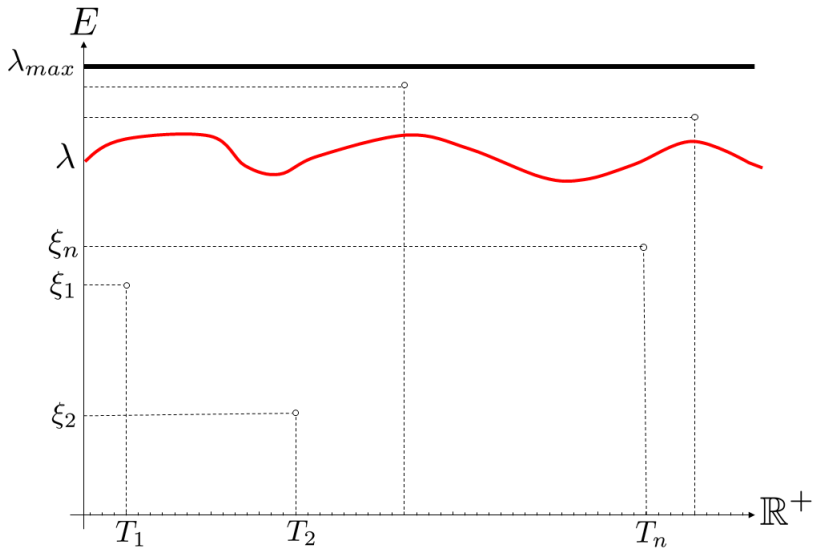


Figure 6.7: Visual representation of the process (6.8). Of all the pairs (T_n, ξ_n) of the Poisson random measure, only those under the function $\lambda(\cdot)$ (in red) are estimated by the function K, those above are ignored. In this way, through an algorithm of acceptance and rejection, the homogeneous Poisson process with intensity $\nu(E) = \lambda_{max}$ becomes non-homogeneous with intensity $\lambda(\cdot)$.

Process with marks that select an event Let X_t be a process that, with at its base a homogeneous Poisson process of intensity λ, at each jump does not have unit increments, but has D possible outcomes: with probability p_i it has a jump of amplitude f_i (possibly dependent on the process itself, $f_i = f_i(x)$), for i from 1 to D (of course $\sum p_i = 1$). We can simulate this process by choosing $E = [0, \lambda]$, $\nu = Leb_{\mathbb{R}}$ and dividing the space E in D disjoint intervals I_i, each of measure λp_i. Then we define $K(x, \xi) = \sum f_i(x) 1_{I_1}(\xi)$. Every time a jump occurs (at time T_n), we randomly extract the mark ξ_n, uniformly distributed in $E = [0, \lambda]$. If this mark falls in the interval I_i (and it occurs with probability p_i) then K will be worth the desired quantity f_i (see Figure 6.8).

$$X_t - X_0 = \int_0^t \int_0^\lambda \sum_{i=1}^{D} f_i(X_{s^-}) 1_{I_1}(\xi) \mathcal{N}(ds, d\xi) \tag{6.9}$$

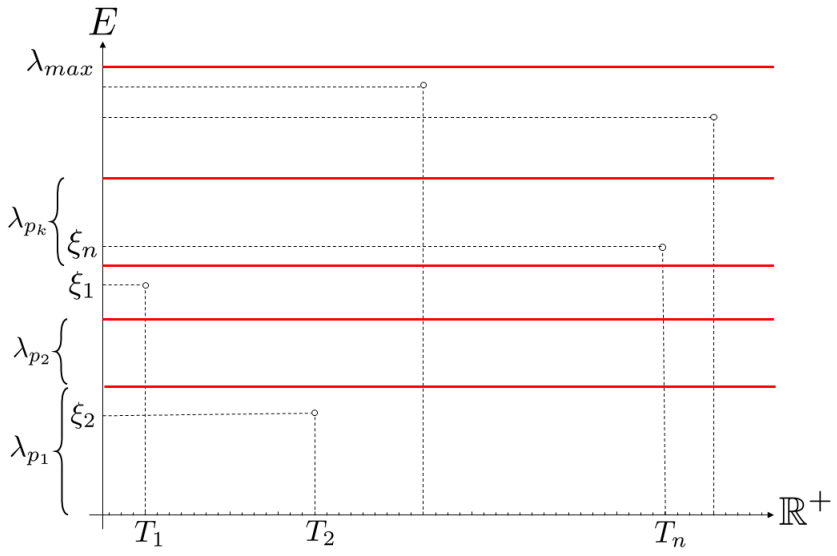


Figure 6.8: Visual representation of the process (6.9). The space E is divided in D disjoint intervals I_i and the function K evaluate the pairs $(T_n, \xi_n) \in \mathbb{R}^+ \times E$ based on their position on the y axis. If ξ_n is in the interval with length λ_{p_k}, hen there will be a jump with amplitude f_k. In this case, for example, the second pair (T_2, ξ_2) produces a jump with amplitude f_1.

Single process, sum of homogeneous Poisson processes. Let $\left(N_t^i\right)_{t\geq 0}$, $i = 1, \ldots, D$ be independent homogeneous Poisson processes, each with intensity λ_i. The process X_t has jumps of a constant amplitude γ_i whenever the N_t^i process has a jump (with the necessary modifications, this is the basic structure of the rare-mutations model [4], as shown in Chapter 5, Section 5.2); it is possible to write the process as the sum of jump processes in the form presented above:

$$X_t = X_0 + \sum_{i=1}^{D} \int_0^t \int_{E_i} \gamma_i \mathcal{N}\left(ds, d\xi_i\right)$$

You can then use the mark space to reformulate the process and write an equivalent one with the same infinitesimal generator that contains only one random measure. In fact, we define, $\lambda_{max} = \sum \lambda_i$; $E = [0, \lambda_{max}]$, with Lebesgue measure. The random measure associated with $\mathbb{R}\times E$ generates a Poisson process N_t^{max} with intensity given by the sum of the intensity of the individual processes. The marks are used to select, after a jump of N_t^{max}, which of the N^i processes has to be considered. As above, defining $K(\xi) = \sum \gamma_i 1_{I_i}(\xi)$, with I_i of measure λ_i, partition of $[0, \lambda_{max}]$, the process takes the form:

$$X_t = X_0 + \int_0^t \int_E K\left(\xi\right) \mathcal{N}\left(ds, d\xi\right) = X_0 + \int_0^t \int_{[0,\lambda_{max}]} \sum_{i=1}^{D} \gamma_i 1_{I_i}\left(\xi\right) \mathcal{N}\left(ds, d\xi\right).$$

6.7.2 Markov Chains as jump processes

An additional difficulty occurs when there is the need to construct a process in which the jumps are a function of the "state before the jump", in which both the frequency and the set

of possible outcomes of the jump are exclusively dependent on the previous state. We use to deal with this a discrete states continuous Markov chain, where for every state there is a coefficient of permanence in the state, and a set of states reachable with relative probability. Here below we write the explicit expression of a pure jump process that is a Markov chain. This expression will be the basis of the complete process.

Fact. *[84] Let X_t be a Markov chain, continuous, with space state $M = \mathbb{Z}$ and transition matrix $Q = (q_{ij})_{i,j\in M}$. Define the function $K : M \times \mathbb{R} \longmapsto \mathbb{R}$ as*

$$K(i,\xi) = \sum_{j\in M} (j-i)\, 1_{\Gamma_{ij}}(\xi),$$

then the continuous time Markov chain X_t satisfies the SDE:

$$X_t = X_0 + \int_0^t \int_{\mathbb{R}} K(X_s,\xi)\mathcal{N}(ds,d\xi)$$

where $\mathcal{N}$ is a Poisson random measure with intensity measure $dtd\xi$.

We recall that row i contains all the information needed for the state i:

- $\sum_{j\in M} q_{ij} = 0$; $q_{ij} \geq 0$ for $j \neq i$ and $-q_{ii} = \sum_{i\neq j} q_{ij}$;
- $q_i := -q_{ii}$ represents the average time of permanence in state i, that is an exponential random variable with intensity q_i;
- q_{ij} for $j \neq i$ are proportional to the jump probabilities, $\mathbb{P}(X_i \longmapsto X_j) = \frac{q_{ij}}{-q_{ii}}$.

We suppose that $\sum_{i\in M} q_i = \infty$. The idea is to use mark space $(E,\nu) = (\mathbb{R}, Leb_{\mathbb{R}})$ (of infinite measure) and divide it into finite-dimensional subspaces q_i as done previously. Each one will be related to a different state. On each subspace we consider a Poisson process with intensity q_i, independent of each other. The intensity function of the jump K, function of $(i,\xi) \in M\times\mathbb{R}$, is non null only when ξ belongs to the subspace related to status i, the one of measure q_i. More in detail:

Let Γ_{ij} be defined as:

$$\Gamma_{12} = [0, q_{12}), \Gamma_{13} = [q_{12}, q_{12}+q_{13}), \ldots$$

$$\Gamma_{21} = [q_1, q_1+q_{21}), \Gamma_{23} = [q_1+q_{21}, q_1+q_{21}+q_{23}), \ldots$$

$$\vdots$$

$$\Gamma_{n1} = \left[\sum_{i=1}^{n-1} q_i, \sum_{i=1}^{n-1} q_i + q_{n1}\right), \ldots$$

Let the sets Γ_i be defined as

$$\Gamma_i = \bigcup_{j\neq i} \Gamma_{ij}.$$

Note that the intervals Γ_i measures q_i. The intervals Γ_{ij} are disjoint and their union gives $\mathbb{R}$, i.e.,

$$\Gamma_{ij} \cap \Gamma_{kl} = \emptyset \text{ if } (i,j) \neq (k,l)$$

$$\bigcup_{\substack{i,j \in M \\ i \neq j}} \Gamma_{ij} = \bigcup_{i \in M} \Gamma_i = \mathbb{R}$$

Also, define the function $K : M \times \mathbb{R} \longmapsto \mathbb{R}$ as

$$K(i,\xi) = \sum_{j \in M} (j-i)\, \mathbb{I}_{\Gamma_{ij}}(\xi), \tag{6.10}$$

then the continuous time Markov chain X_t satisfies the SDE:

$$X_t = X_0 + \int_0^t \int_{\mathbb{R}} K(X_s, \xi)\, \mathcal{N}(ds, d\xi)$$

where $\mathcal{N}$ is a Poisson random measure with intensity measure $dt \times d\xi$.

The variable ξ lives on the real line, with Lebesgue measure, therefore E has infinite measure. For this reason we decide to divide the real line in the disjoint intervals $\{\Gamma_i\}_{i \in \mathbb{N}}$, measuring $m(\Gamma_i) = q_i$. The random measure will therefore consist of the sum of the measures $\mathcal{N}_i$ on Γ_i, each a sequence of pairs $\left(T_n^i, \xi_n^i\right)$, with $n \in \mathbb{N}$ and every i related to a state in M. The times $\left(T_n^i\right)_{n \in \mathbb{N}}$ are the arrival times of the homogeneous Poisson process N^i with intensity q_i, independent of any other process N^j. The random variables $\left(\xi_n^i\right)_{n \in \mathbb{N}}$ are uniform on Γ_i, thus they belong with probability $\frac{m(\Gamma_{ij})}{m(\Gamma_i)}$ to the interval Γ_{ij}. As long as the process is in the state $j \in M$, the function K is not null in Γ_j, therefore it will ignore all the other infinite Poisson processes which, at the same time, occur on Γ_i, $i \neq j$, obtaining all the characteristics of a Markov chain (see Figure 6.9 for a visual representation of this costruction).

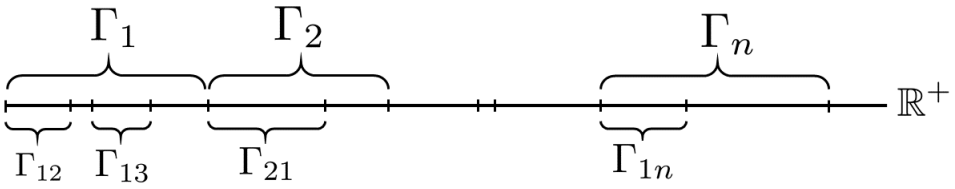


Figure 6.9: Visual representation of the space $E = \mathbb{R}^+$ for the construction presented above for a continuous time Markov chain. Since the measure of space E is infinite, the construction presented in (6.9) is repeated for each interval Γ_j. For each interval Γ_j a Poisson process is generated, independent from any other, and the function K (6.10) will select jumps in Γ_j only if the current state of the process is $j \in \mathbb{N}$.

6.8 Mutation process

After the previous Section we have all the tools to build the model of mutations (pure jump) in $\mathcal{S}_{\mathbb{R}}^{\infty}$, i.e. to complete the model (6.7):

$$N_t = N_0 + \int_0^t \tilde{a}\left(N_s\right) ds.$$

This model takes into account the selection and adaptation mechanism, acting on the phenotypes. We want to add a stochastic process to (6.7), obtaining:

$$N_t = N_0 + \int_0^t \tilde{a}\left(N_s\right) ds, + \int_0^t \int_E K(N_{s^-}, \xi)\mathcal{N}\left(ds, d\xi\right).$$

The underlying biological process is the following: the number of individuals in a population divided in D different initial genotypes changes only according to selection and adaptation (the deterministic term given by $\tilde{a}$), until a mutation happens. A mutation occurs randomly in the population, with a temporal frequency proportional to the total number of individuals, and transform a small number of them with the same genotype (1 to 3) into new individuals, with a similar, but different, genotype. After the mutation, the process increase in dimensionality (there are now $D+1$ different genotypes), and continue to evolve deterministically until the next mutation. In detail, we are trying to build, with a logic similar to the one used for Markov chains in Section 6.7, a jump process with space (E, ν) and an appropriate function K. We require that the mark space E takes into account the following phenomena:

- the temporal frequency of the process, that has to be function of the total number of individuals;
- identification of the individuals (all with the same genome) that suffer a mutation;
- choice, in proportion to the type of genome identified, of the type of genome that the newborns express.

In what follows we construct the space E, with elements ξ, with measure ν and define the function K.

Mutation event and identification of the mutants

A mutation event occurs. The process is based on a non-homogeneous Poisson of intensity proportional to the number of individuals in the population $\lambda = \lambda(N)$, with $\lambda(N)$ positive and bounded by λ_{max}. The jumps happen therefore with frequency proportional to the number of individuals (with constant population the Poisson is homogeneous). To take into account this type of temporal frequency of the process, we define the first variable of the mark space E as $z \in \mathbb{R}$ (with Lebesgue measure) and we write:

$$K\left(N, \xi\right) = 1_{[0,\lambda(N)]}(z)\bar{K}\left(N, \xi\right)$$

We divide the set $\mathbb{R}$ into finite intervals, one of which is $[0, \lambda_{max}]$, and we ignore everything that happens outside of this. With an acceptance-rejection mechanism, we have the target intensity of the Poisson process.

Identification of the mutants. Mutations occur "blindly" by randomly selecting a small group of individuals with a fixed genome. The choice is proportional to the frequency of the different types in the population at the time of the mutation. To write this in the model we consider the unit interval $[0,1]$. Then we consider the family of sets $\{I_i\}_{i\in\mathbb{N}}$, a partition of $[0,1]$ ($\bigcup I_i = [0,1]$, $I_i \cap I_j = \emptyset$ for $i \neq j$), such that $Leb(I_i) = \frac{n_i}{\sum n_j}$. For each genome then we associate a set with measure equal to its frequency. It should be noted that only a finite number of subsets has non-null measure, effectively dividing the interval into a finite number of parts. Calling u a variable that lives in $[0,1]$, the function $\bar{K}$ it is defined as follows:

$$\bar{K}(N,\xi) = \sum_{i\in\mathbb{N}} \{1_{I_i}(u) K_i(N,\xi)\};$$

Choice of the new genome

Once the genome that changes has been selected, the new genome must be similar to it. To achieve this we use the construction presented above in Section 6.7 for Markov chains.

"There are many more possible proteins than available protons [in the Universe] and hence evolution will explore a vanishingly small subset of all possible proteins. What is true of proteins is also true of genes and genomes. We can imagine all nucleotide sequences arranged in a way the nearest neighbors differ in one position. Evolution is a trajectory through sequence space [73]".

First, it is necessary to define a concept of "distance" between the genomes, in terms of mutations: a genome explores space through mutations, which transform it into a new element of $\mathcal{G}$, which is "near" him, in some sense; on $\mathcal{G}\times\mathcal{G}$ we define a weighted graph. If we imagine to follow the evolutionary history of a genome, we would see it moving between one vertex and another, following the transition probabilities of a Markov chain:

Definition 29. Let G_t be a discrete Markov chain on state space $(\mathcal{G}, \mathcal{P}(\mathcal{G}), \mathbb{P})$. On the cartesian product $\mathcal{G}\times\mathcal{G}$ is then defined a jump matrix Π, such that:

$$G_t = \text{genome } g \in \mathcal{G} \text{ at time } t;$$

$$\Pi_{ij} := \mathbb{P}(Y_{t+1} = j | Y_t = i).$$

Π is an irreducible matrix.

The graph of the Markov chain, representation of the jump matrix, makes evident a measure of closeness between one genome and another. The matrix Π has been chosen irreducible, so that it is always possible, with a sufficient number of mutations, to move from one genotype to another. The choice is in line with the evolutionary theory, which presupposes all the species descending from a single common genome, as long as we suppose the single mutational events reversible.

As said above, we want to use a construction similar to the one in Section 6.7 for Markov chains; in that case the space $\mathbb{R}$ was divided into subsets, each related to a state of the chain. When the process was in the state i, the function K allowed the variable ξ to move only in the subset relative to i, and to select the value of the jump according to the row i of the transition matrix. Unlike that construction, the space to be divided here is not the real line, but the Cartesian product $\mathcal{G}\times\mathcal{G}$. We impose that when the mutation affects the genome g,

the function K allows the variable ξ to move only in the subset relative to the state g, that is $\{g\} \times \mathcal{G}$. On this space, thanks to the jump matrix Π we define μ_g, which measures the probability that g has to change into other genomes.

Definition 30. Fixed $g \in \mathcal{G}$, the measure $\eta_g : \mathcal{G} \longmapsto \mathbb{R}^+$ is defined as $\eta_g(h) := \Pi_{g.h}$. η_g is a probability measure on $\mathcal{G}$, for all $g \in \mathcal{G}$.

On $\mathcal{G} \times \mathcal{G}$ is therefore defined the measure η:

Definition 31. η is a measure,$\eta : \mathcal{G} \times \mathcal{G} \longmapsto \mathbb{R}^+$, such that $\eta((g,h)) := \eta_g(h)$, i.e. $\eta|_{\{g\}\times\mathcal{G}} = \eta_g$. The measure η clearly is not a probability measure, because $\eta(\mathcal{G} \times \mathcal{G}) = \sum_{g\in\mathcal{G}} \eta_g(\mathcal{G}) = \infty$.

See Figure 6.10.

Example. In Nature, indicating a genome as a string of finite length of letters $\{A, C, T, G\}$, i.e. the space of potential genomes

$$\mathcal{G} = \bigcup_{n\in\mathbb{N}} \{A, C, T, G\}^n ;$$

we can define the jump matrix Π thanks to the Hamming distance [40];

Definition 32. The Hamming distance d_H between two strings of equal length is the number of positions in which the corresponding symbols are different. In other words, the Hamming distance measures the number of substitutions necessary to convert one string into another, or the minimum number of mutations that have led to the transformation of one genome into another.

By simplifying, if we imagine that only point mutations exist, and that no more than one mutation can occur simultaneously, then a string of DNA can only change into another that has Hamming distance 1 from it. For example AAA can only change into a string that contains A in two positions out of three. Fixed the genome $g_i \in \mathcal{G}$:

$$\eta_i(g_j) = \Pi_{ij} = \begin{cases} \frac{1}{3L_j} & \text{if } d_H(g_j, g_i) = 1 \\ 0 & \text{otherwise} \end{cases}$$

where 3 is the number of genomes in which g can be transformed by changing one nucleotide and L_j is the length of the genome g_j.

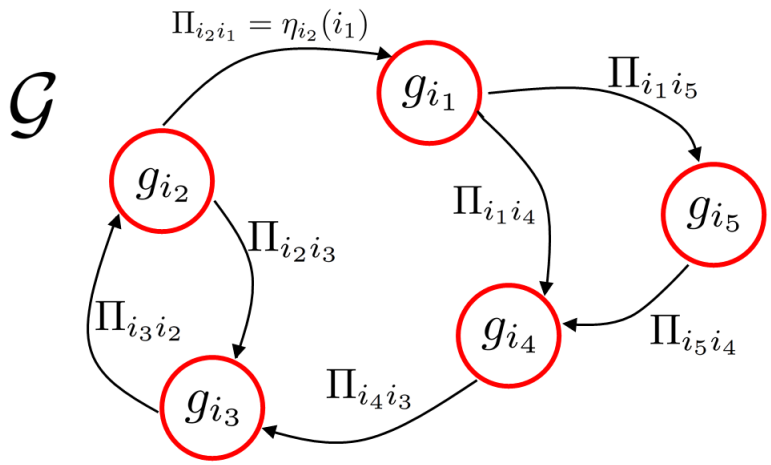


Figure 6.10: Example of graph on the genome space $\mathcal{G}$. The arrows have the weights Π_{jk} and represents mutation jumps. Note that $\Pi_{jk} = \eta_j(k)$.

In line with above, it is possible to define K_i, function of the variable v on $\mathcal{G} \times \mathcal{G}$:

$$K_i(N, \xi) = 1_{\{\alpha(i)\} \times \mathcal{G}}(v) \gamma_i(N, \xi)$$

With γ_i we indicate the intensity of the jump. The amplitude of the jump is r, a very small value (1 to 3), if there are at least r individuals to change; otherwise they all change. With e_i we indicate the element of the canonical base of $\mathcal{S}_{\mathbb{R}}^{\infty}$, $e_i(k) := \delta_{ik}$, $k \in \mathbb{N}$. The jump moves from individuals with genome $g_i = \alpha(i)$ to individuals of the new type v_2, where $v = (\alpha(i), v_2)$, that is

$$\gamma_i(N, \xi) = \min(n_i, r) \left(e_{\alpha^{-1}(v_2)} - e_i\right).$$

See Figure 6.11 for a visual representation of the mark space $E = (\mathbb{R}^+ \times [0, 1] \times (\mathcal{G} \times \mathcal{G}))$.

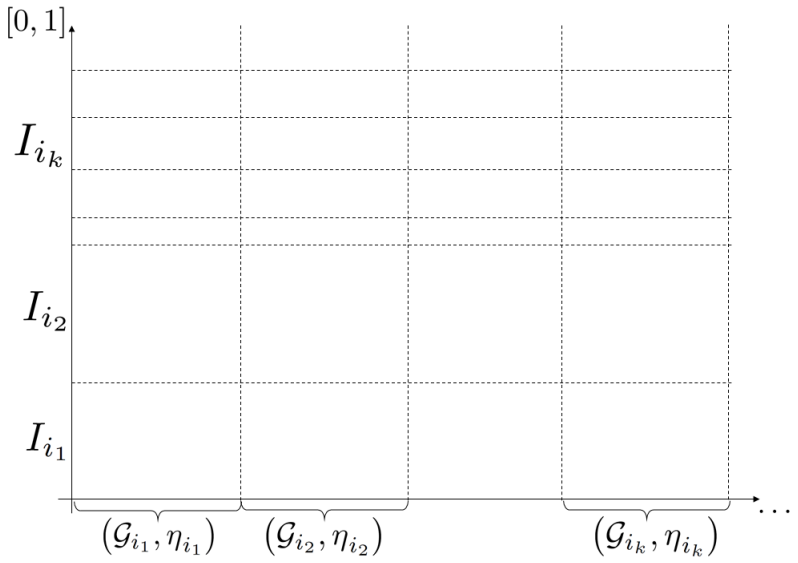


Figure 6.11: Visual representation of the mark space $E = (\mathbb{R}^+ \times [0,1] \times (\mathcal{G} \times \mathcal{G}))$ for the complete model, with function K equal to $K(N,\xi) = 1_{[0,\lambda(N)]}(z) \sum_{i\in\mathbb{N}} \left\{1_{I_i}(u) 1_{\{i\}\times\mathcal{G}}(v) \cdot \min\left(n^i, r\right)\left(e_{\alpha^{-1}(v_2)} - e_i\right)\right\}$. The variable $z \in \mathbb{R}^+$ (not shown) makes the process non-homogeneous and of the desired intensity, as in equation (6.8). The pair of variables (u,v), living in the space shown in the figure ($u \in [0,1]$, on the axis y, $v \in \mathcal{G}\times\mathcal{G}$ on the axis x), regulate the choice of the genome that will change (u), and evaluate which one will change (v). Note the parallelism with the construction for Markov chains presented in Section 6.7, where the set E, of infinite measure, is divided into subsets of finite measure ($(\mathcal{G}, \eta_i)$ here, Γ_k for a Markov chain)

Complete jump process. The mutation process has then the form:

$$N_t = N_0 + \int_0^t \int_E K(N_{s^-}, \xi)\, \mathcal{N}(ds, d\xi);$$

defining, for ease of notation, $\gamma_k\left(n^k, v\right) := \min\left(n^k, r\right)\left(e_{\alpha^{-1}(v_2)} - e_k\right)$, K has the following form:

$$K(N,\xi) = 1_{[0,\lambda(N)]}(z) \sum_{k\in\mathbb{N}} \left\{1_{I_k(N)}(u) 1_{\{\alpha(k)\}\times\mathcal{G}}(v) \gamma_k\left(n^k, v\right)\right\}$$

$$\xi = (z,u,v) \in E = \mathbb{R}^+ \times [0,1] \times (\mathcal{G}\times\mathcal{G})$$

$$\{I_i(N)\}_{i\in\mathbb{N}} \text{ partition of } [0,1],\ m(I_i(N)) = \frac{n^i}{\sum n^j}.$$

$$\nu(E) = \nu\left(\mathbb{R}^+ \times [0,1] \times (\mathcal{G}\times\mathcal{G})\right) = +\infty.$$

To give an intuitive explanation, we follow the construction for Poisson random measures presented in Section 6.7 ([50]). Let $(T_n)_{n\in\mathbb{N}}$ be a homogeneous Poisson process with intensity $\lambda_{max} := \|\lambda\|_\infty$, let $(z_n)_{n\in\mathbb{N}}$, $(u_n)_{n\in\mathbb{N}}$ $(v_n)_{n\in\mathbb{N}}$ be collections of independent random variables,

$$z_n \sim Uniform([0, \lambda_{max}]),$$

$$u_n \sim Uniform\left([0,1]\right),$$
$$v_n \sim Uniform\left(\mathcal{G}\times\mathcal{G}\right);$$

then the process can be rewritten as:

$$N_t = N_0 + \sum_{n:T_n\le t} 1_{\left[0,\lambda\left(N_{T_n^-}\right)\right]}(z_n)\sum_{j\in\mathbb{N}} 1_{I_j\left(N_{T_n^-}\right)}(u_n)1_{\{\alpha(j)\}\times\mathcal{G}}\gamma_j\left(n^j_{T_n^-}, v_n\right). \tag{6.11}$$

This expression presents only random variables independent of each other, and we can give a clearer idea of how the process works. Intuitively, the pure jump process:

- calculates a sequence of times $(T_n)_{n\in\mathbb{N}}$ of a homogeneous Poisson process with intensity λ_{max};
- at each occurrence of a jump T_n calculates z_n, a uniform variable uniform in $[0,\lambda_{max}]$ and accepts it if $z_n \le \lambda\left(N_{T_n^-}\right)$;
  - the first two steps and the succession of random variables z_n therefore are used to generate a process with a non-homogeneous intensity dependent on the process itself $\lambda(\cdot)$
- if the jump is accepted calculates a variable $u_n \sim Unif\left(0,1\right)$ and calculates j such that $u_n \in I_j\left(N_{T_n^-}\right)$;
  - variables u_n select the type of genome that will undergo a mutation;
- only at this point the variable v_n is generated, on which calculate the function γ_j.
  - Note that, fixed j, the function K is non null only if $v_n \in \{\alpha(j)\}\times\mathcal{G}$, so, instead of v_n we can imagine a variable v_n^j on $v_n \in \{\alpha(j)\}\times\mathcal{G}$ is generated, using η_j as measure
  - variables v_n select, starting from the genome that is changing, a new born similar to the predecessor (according to the measure η), which will be generated in small quantities.

6.9 Complete process, existence and uniqueness

Let us now write the complete equations for the model and rely on [7] for what regards the existence and uniqueness of the process.

Consider the stochastic equation

$$\begin{cases} N_t = N_0 + \int_0^t \tilde{a}\left(N_s\right)ds + \int_0^t\int_E K\left(N_{s^-},\xi\right)\mathcal{N}\left(ds,d\xi\right), & t>0 \\ N_0 \in \mathcal{S}^\infty_{\mathbb{R}}, \end{cases} \tag{6.12}$$

in $\mathcal{S}_{\mathbb{R}}^{\infty}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a right-continuous filtration $\{\mathcal{F}_t,\, t \geq 0\}$, $\tilde{a}$ and K Borel-measurable functions of appropriate dimensions:

$$\tilde{a}\,(N) = \begin{cases} \frac{F(N)}{\sum_{vit}} n^k + n^k \left(\Phi_k(N) - \frac{1}{\sum_{vit}} \sum_{j:\, g_j \in \mathcal{G}-\mathcal{G}_0} n^j \Phi_j(N) \right), & k:\, g_k \in \mathcal{G} - \mathcal{G}_0 \\ -\lambda n^k,\ \lambda \gg 0, & k:\, g_k \in \mathcal{G}_0 \end{cases} \tag{6.13}$$

where $\Sigma_{vit} = \sum_{j:\, g_j \in \mathcal{G}-\mathcal{G}_0} n^j$, F equation of a logistic growth with carrying capacity K_F and r_F rate of increase,

$$F\,(N) = r_F \left(1 - \frac{\sum_{j:\, g_j \in \mathcal{G}-\mathcal{G}_0} n^j}{K_F} \right) \sum_{j:\, g_j \in \mathcal{G}-\mathcal{G}_0} n^j;$$

$(\Phi_k)_{k \in \mathbb{N}}$ fitness functions with

$$\sup_{j:\, g_j \in \mathcal{G}-\mathcal{G}_0} |\Phi_k\,(N)| = L < \infty;$$

The jump term K is

$$K\,(N, \xi) = 1_{[0, \lambda(N)]}(z) \sum_{i \in \mathbb{N}} \left\{ 1_{I_i}(u) 1_{\{i\} \times \mathcal{G}}(v) \cdot \min\left(n^i, r\right) \left(e_{\alpha^{-1}(v_2)} - e_i \right) \right\}. \tag{6.14}$$

Remark. A SOLUTION of the process presented above is a cadlag process with values in $\mathcal{S}_{\mathbb{R}}^{\infty}$ defined on the probability space, such that N_t is $\mathcal{F}_t-$adapted and there exist an $\mathcal{F}_t-$stationary Poisson process having Poisson random measure $\mathcal{N}(\cdot, \cdot)$ on $\mathbb{R}^+ \times E$,

$$E = \left(\mathbb{R}^+ \times [0,1] \times (\mathcal{G} \times \mathcal{G})\right),$$
$$\xi = (z, u, v)\,,$$

with mean measure $m \times \nu$:

$$\nu = Leb_{\mathbb{R}^+} \times Leb_{[0,1]} \times \eta,$$

ν $\sigma-$finite measure on a measurable space $(E, \mathcal{E})$ such that (6.12) holds a.s. and $N_0 \in \mathcal{F}_0$ is independent of the Poisson increments.

Theorem 33. *Under the assumption of the above remark, a solution of (6.12)(6.13)(6.14) exists and is unique for all $t \geq 0$.*

Proof. If there are no jumps, i.e. the process is $\tilde{N}_t$ (6.12), with $K = 0$, is deterministic and is set on $\mathbb{R}^D$, where D is the number of non-zero elements in the support of N_0.

$$n_t^k = \begin{cases} n_0^k + \int_0^t \tilde{a}_k\,(N_s)\, ds, & \text{if } n_0^k \text{ is one of the } D \text{ non-null elements of } N_0 \\ 0, & \text{otherwise} \end{cases}$$

There are no way, except for mutations, to increase (or decrease) the dimension of the problem. Then, we can use classical arguments to guarantee existence and uniqueness for the solution of:

$$\begin{cases} \dot{n^k} = \frac{F(n^1, \ldots, n^d)}{\sum_{k=1}^d n^j} n^k + n^k \left(\Phi_k(n^1, \ldots, n^d) - \frac{1}{\sum_{k=1}^d n^j} \sum_{j=1}^d n^j \Phi_j(n^1, \ldots, n^d) \right) & k = 1, \ldots, d \\ \dot{n^{0_j}} = -\lambda n^{0_j}, & j = 1, \ldots, v \end{cases}$$

with $d + v = D$, $\lambda \gg 0$, $N_0 \in \mathcal{S}_{\mathbb{R}}^{\infty}$, i.e. $N_0 \in \mathbb{R}^D$, $n_0^k > 0$ for $k = 1, \ldots D$,

- $\tilde{a} : \Omega \subseteq \mathbb{R} \times \mathbb{R}^D \longmapsto \mathbb{R}^D$ is smooth near the initial datum $(t_0, N_0) = (0, N_0)$, so it is locally lipschitz in the open set $\Omega \ni (0, N_0)$, then the solution exists and is unique locally;
- $N_{tot} := \sum_{k=1}^{D} n^k$ is such that $\dot{N}_{tot} = F(N_+) - \lambda \sum_{j=1}^{\nu} n^{0_j}$, with F logistic, so $N_{tot} \in [0, K]$ for all $t \geq 0$ when $n_0^k \geq 0$ for all k;
- $n^k(t) = 0$, for $t \geq 0$, is a solution; then $n_k(t)$ can not change in sign. Because of, for all k $n_0^k > 0$, then $0 \leq n_t^k \leq K$.
- $\tilde{a}$ is bounded, so the solution exists globally in time.

$$\sum_{k=1}^{d} \left| \frac{F(n^1, \ldots, n^d)}{\sum_{j=1}^{d} n^j} n^k + n^k \left(\Phi_k(n^1, \ldots, n^d) - \frac{1}{\sum_{j=1}^{d} n^j} \sum_{j=1}^{d} n^j \Phi_j(n^1, \ldots, n^d) \right) \right| + \sum_{j=1}^{v} \left| -\lambda n^{0_j} \right|$$

$$\leq \sum_{k=1}^{d} n^k \left\{ r_F \left| \left(1 - \frac{\sum_{j=1}^{d} n^j}{K_F} \right) \right| + |\Phi_k| + \frac{1}{\sum_{j=1}^{d} n^j} \sum_{j=1}^{d} n^j \left| f^j \right| \right\} + \lambda \sum_{j=1}^{v} n^{0_j}$$

$$\leq \sum_{k=1}^{d} n^k \left\{ r_F + L + \frac{\max_{N \in \mathbb{R}^D} |\Phi_j|}{\sum_{j=1}^{d} n^j} \sum_{j=1}^{d} n^j \right\} + \lambda \sum_{j=1}^{v} n^{0_j}$$

$$\leq \sum_{k=1}^{d} n^k \left\{ r_F + 2L \right\} + \lambda \sum_{j=1}^{v} n^{0_j} \leq \max \left\{ r_F + 2L;\, \lambda \right\} K_F.$$

Next, we use the results in [7]. If we define the function Λ:

$$\Lambda(N) := \nu \left\{ \xi \in E,\, |K(N, \xi)| \neq 0 \right\},$$

$$\Lambda(N) := \nu \left\{ (z, u, v) \in E, \left| 1_{[0, \lambda(N)]}(z) \sum_{i \in \mathbb{N}} \left\{ 1_{I_i}(u) 1_{\{\alpha(i)\} \times \mathcal{G}}(v) \cdot \min\left(n^i, r\right) \left(e_{\alpha^{-1}(v_2)} - e_i \right) \right\} \right| \neq 0 \right\},$$

we can show that $\Lambda(N) < \infty$: the amplitude of the jump $\gamma_i(N, \xi) = \min\left(n^i, r\right) \left(e_{\alpha^{-1}(v_2)} - e_i \right)$ is null only when n^i is null, so the function K is null only when $z \in [0, \lambda(N)]$, and for all $i \in \mathbb{N}$ $u \in I_i$, $\nu \in \{\alpha(i)\} \times \mathcal{G}$, then

$$\Lambda(N) = \nu \left\{ [0, \lambda(N)] \times \bigcup_{i \in \mathbb{N}} I_i \times (\{\alpha(i)\} \times \mathcal{G}) \right\} \leq \nu \left\{ [0, \lambda_{max}] \times \bigcup_{i \in \mathbb{N}} I_i \times (\{\alpha(i)\} \times \mathcal{G}) \right\}$$

$$= \lambda_{max} \cdot \sum_{i \in \mathbb{N}} m(I_i)\, \eta(\{\alpha(i)\} \times \mathcal{G}) = \lambda_{max} \cdot \sum_{i \in \mathbb{N}} m(I_i) = \lambda_{max} < \infty.$$

These calculations guarantee that $\tau_1 = \inf\left\{ t > 0, \int_0^t \int_E K\left(\tilde{N}_{s^-}, \xi \right) \mathcal{N}(ds, d\xi) \neq 0 \right\}$ fulfills $\mathbb{P}(\tau_1 > 0) = 1$ for all initial positions $N_0 \in \mathcal{S}_{\mathbb{R}}^{\infty}$ [7]. Starting at N_0, N_t evolves as $\tilde{N}_t$ for $t < \tau_1$ and then jumps to

$$N_{\tau_1} = \tilde{N}_{\tau_1} + K(\tilde{N}_{\tau_1}, \xi_1),$$

where (τ_1, ξ_1) is a point belonging to the Poisson random measure $\mathcal{N}(\cdot,\cdot)$. Next, starting at N_{τ_1}, the process evolves as $\tilde{N}_t$ (with $\tilde{N}_0 = N_{\tau_1}$). Now, after the first jump, if the mutation has created a new genome, the deterministic process is on $\mathbb{R}^{D+1}$, and we can prove again existence and uniqueness with the above assumptions. The process evolves up to a random time τ_2 and so on. We can define the jump times $\tau_2, \tau_3, \ldots$ appropriately and the process N_t is well defined for all $t < \tau_\infty := \lim_{n\to\infty} \tau_n$. Because of

$$\sup\{N \in \mathcal{S}^\infty_{\mathbb{R}},\, \Lambda(N)\} = \lambda_{max} < \infty,$$

then $\tau_\infty = \infty$ with probability 1 for all initial conditions [7]. Then global existence and uniqueness are verified. □

Conclusions

In this chapter we built a population dynamics model of selection, adaptation and mutation that generalizes the replicator dynamics with rare mutations presented in [4, 5] (see Chapters 4 and 5). The model is structured to follow the evolution over time of a population of different types of potentially infinite genotypes. For this reason the setting has shifted from the simplex $\mathcal{S}^D$ of the replicator dynamics to a space of infinite size, $\mathcal{S}^\infty_{\mathbb{R}}$. The genotypes are subject to mutations, which occur randomly in the population, without favoring one type over another, and generating individuals with features never seen before, exploring with time the infinite set of potential genotypes. We have therefore constructed a stochastic process that takes these characteristics into account. Each genotype expresses a given phenotype, or a set of characteristics that are subject to natural selection. After defining the GP functions, which map genotypes into phenotypes, we have written the equations that regulate the selection and adaptation of these phenotypes. The model thus obtained was written in a formal manner and its existence and uniqueness were discussed. After this chapter of model setup, in the next we analyze some examples, without specific biological relevance, which highlight particular dynamic characteristics of the process.

Chapter 7

Simulations of the Genotype-Phenotype model

7.1 Numerical implementation

We present in this chapter various examples of applications of the model in the previous chapter, see Section 6.9. Using the intuitive interpretation of the mutation process given in Section 6.8, we implement the model in the following way:

- We define the initial parameters, such as the final time T, the time step Δt, the mortality rate of individuals with non-vital phenotype, the maximum intensity λ_{max} of the Poisson process for mutations.

  The time step of the simulation, necessary to approximate the non-stochastic differential equations, is chosen so that within it the probability to have *at most* a mutational jump is almost one (is above a threshold chosen a priori).

- The homogeneous Poisson process of intensity λ_{max} is generated until time T and the succession of jump times T_n is saved;

- We define the initial data by choosing a finite number D of genomes that are in the population; D is the actual size of the problem.

  The initial number of individuals N_0 is defined here, which is split into D different types, with relative genotypes and phenotypes in a vector with D components, finite. Although in theory the problem is on $\mathcal{S}_{\mathbb{R}}^{\infty}$, the simulation considers a finite vector, on $\mathbb{R}^D$.

The iterations start. For each time interval:

- A mutation occurs if in the considered interval there is a jump of the homogeneous Poisson process; if so (let $\bar{t}$ be the time) an acceptance-rejection technique is carried out to check if this jump can be considered.

  The homogeneous random variable z_n is sampled with uniform distribution on $[0, \lambda_{max}]$; if this number is less than or equal to $\lambda(N_{\bar{t}-})$, then the jump is accepted and a mutation occurs.

- If a mutation occurs:

- A random genome g_k, $k = 1, \dots, D$ is chosen in the population, proportionally to the amount of individuals with that genome, using the random variable $u_n \sim Unif(0,1)$: the higher is n^k, the higher is the probability to be chosen.
- A new genome is generated starting from the chosen one; the function responsible for this step takes the parent's genome as input and outputs the genome obtained by applying a step to the Markov chain with transition matrix Π. The new born is labeled as the $(D+1)$th genome and the problem changes its actual size, increasing by one.

  It is important to note that the matrix Π may not be known a priori, as the function of the program provides an algorithm to modify the initial genome randomly, obtaining a new one, without explicitly calculating the values of Π. For example, if the genome g_k were a string of nitrogenous bases, the function could act as a "natural" point mutation, choosing one of the random nitrogenous bases, and modifying it, at random. As the new born is labeled as $(D+1)$th genome, the process becomes a vector on $\mathbb{R}^{D+1}$ (the mutation can rarely transform a genome into one already present in the population and not generate a new genome, in which case the problem remains on $\mathbb{R}^D$).

  In theory, the new genome is already present in the population vector N, element of $\mathcal{S}^\infty_{\mathbb{R}}$ in an appropriate position, given a priori from the sorting function α; in the simulation, however, from a finite vector, we obtain a vector of size greater than one, with the new born that occupy the position $D+1$. In this way, with time, the sorting is obtained step by step and represents the temporal order of appearance of a given genome in the population.
- The number of individuals changes according to the law of the process, which eliminates $\min(n^k, r)$ individuals of the type k and creates as many of the new type $D+1$.
- The GP-map of the new genotype is calculated, defining its phenotype.

  The passage between genotype and phenotype is summarized here; the GP-map function is defined for each genotype. Given the nature of the problem considered, this function can be computationally heavy (as in cases where the map is the development of a cellular automaton), but in any case it is calculated only once, when the creation of the given genome happens.

- After the possible mutation, a step in the deterministic process is made, using Heun's method.

  - The phenotypes are divided between vital and non-vital (the latter indicated with $'0'$);
  - The vital phenotypes undergo a step of the replication dynamics with growth function:
    - an intermediate value $\tilde{n}^k_{t+1}$ is calculated using Euler's method:

$$\tilde{n}^k_{t+1} = n^k_t + \Delta t \tilde{a}\left(N^+_t\right);$$

* the final approximation n_{t+1}^k is obtained as:

$$n_{t+1}^k = n_t^k + \frac{\Delta t}{2}\left[\tilde{a}\left(N_t^+\right) + \tilde{a}\left(\tilde{N}_t^+\right)\right];$$

– Non-vital phenotypes decrease exponentially, $n_{t+1}^k = n_t^k e^{-\lambda \Delta t}$;

It should be noted that the fitness function can not take genotypes as input, thus being invariant for individuals in the same neutral space.

- The variables of the problem are updated, including the vector containing the number of individuals and the vectors containing genotypes and phenotypes present at that time in the population, the actual dimension of the problem, which could be increased.

7.2 Examples of applications

We propose here different examples to highlight the features of the genotype-phenotype model (6.12)(6.13)(6.14). As defined, in the model, the space of the $\mathcal{G}$ genomes, the matrix relative to the mutations Π, the GP-map and the fitness functions Φ, are not specified. When these parameters are changed, different models are obtained; for this reason, for each example we initially detail the choices made.

Fibonacci GP-Map.

A structurally simple example, but able to summarize the main characteristics of the model, is that built from the "Fibonacci GP-map", presented in [36].

Genome space and mutations. The genome space $\mathcal{G}$ is the set of binary strings of fixed length, $\mathcal{G} = \bigcup_{n\in\mathbb{N}} \{0,1\}^n$, a genome therefore represents a simplified DNA string, in which the nitrogenous bases are reduced from four to two, 0 and 1. When a mutation occurs in the genome, two different events may happen:

- with probability $(1-p)$, $p \in (0,1)$ the string undergoes a punctual mutation, i.e. an element chosen randomly within it changes from 0 to 1 or vice versa;
- with probability p instead an insertion occurs, i.e. a 0 or a 1 is added into the genome, in a random position; through insertions, the effective dimensionality of the problem increases by one;

GP-map. A genome g of length L, $g \in \{0,1\}^L \subset \mathcal{G}$ expresses a phenotype $f \in \mathcal{F}$, which is a binary string of length shorter than L, obtained in the following way:

- starting with the first digit the sequence is considered "coding" until a "stop codon" is encountered,
- after the stop codon the sequence is considered "non-coding".

Each possible sequence up to the first occurrence of the stop codon 11 uniquely maps to a distinct phenotype. The sequence after the first stop codon, on the other hand, gives rise to the neutral space of that phenotype. If the stop codon does not exist, then the phenotype obtained is non-vital, and it is indicated with 0.

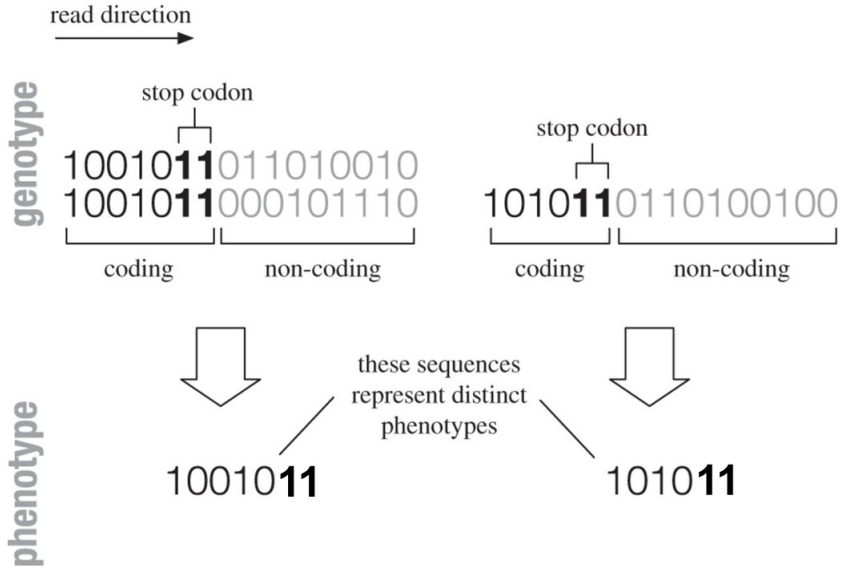


Figure 7.1: Three examples of genotype sequences, which map two different phenotypes in the Fibonacci genotype-phenotype map. Reading from the left the sequence is regarded as "coding" up to the first occurrence of the "stop codon" sequence 11. Thereafter the sequence is regarded as "non-coding". Each possible coding sequence represents a different phenotype, whereas the non-coding sequence leaves the phenotype entirely unaffected. Figure by [36].

Fitness 1. Let us define the fitness matrix of the two-strategist game Hawks and Doves, for $G = 2$, $C = 3$,

$$\mathcal{U} = \begin{pmatrix} \frac{G-C}{2} & G \\ 0 & \frac{G}{2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & 2 \\ 0 & 1 \end{pmatrix};$$

fixed time t, the fitness of phenotype f_k is obtained by playing multiple times the game of matrix $\mathcal{U}$ against the phenotype which has the higher number of individuals in the population at that time, namely $f_{max,t}$, in the following way:

- define $l = \max(\text{length of } f_k, \text{length of } f_{max,t})$
- for i from 1 to l, the two players challenge each other, choosing their strategy based on what they have encoded in position i of their genotype:
  - if position i is equal to 0, the player choose Dove;
  - if position i is equal to 1, the player choose Hawk;
  - if position i is greater than the length of the genome, the player choose Hawk.

  fenotype f_k has a reward of $\phi(i, f_k, f_{max,t})$ each time.

- After the l fights, the result are averaged.

$$\Phi_k(N,t) = \frac{1}{l}\sum_{i=1}^{l}\phi(i, f_k, f_{max,t}).$$

Examples. Suppose that $p = 0$, i.e. that no insertion mutations can occur, only punctual; the initial length of the genomes is $L = 3$, and because of $p = 0$, can not increase. In this situation, only 8 different genotypes can be generated by mutation, with the following genotypes:

g	$\overset{GP}{\longmapsto}$	f
000		0
001		0
010		0
011		011
100		0
101		0
110		11
111		11

The only vital phenotypes are therefore $f_1 = 011$ and $f_2 = 11$; the value 0 represents the strategy "Dove", while the value 1 the strategy "Hawk".

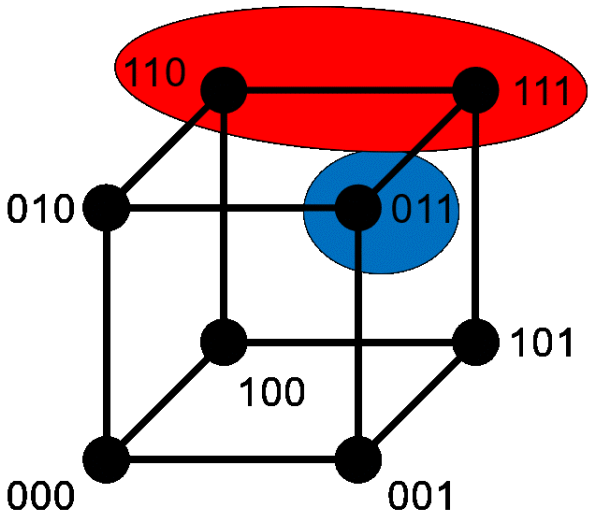


Figure 7.2: Genome space $\mathcal{G}$ with graph given by Π when there are no insertion mutation and $L = 3$. Genomes 111 and 110 are the neutral space of phenotype 11 (in red) while 011 is the only element in the neutral space of phenotype 011; all other genotypes are non-vital (phenotype 0).

In a population initially formed by the majority of individuals of phenotype $f_1 = 011$, fitness are calculated challenging an individual that use the set of strategies 011, i.e. "First Dove, then Hawk, then Hawk":

- *Fitness of phenotype* f_1. Individuals of phenotype f_1 fight a total of 3 times, the first time acting like Doves (0), the second and third acting like Hawks (11).

$$\Phi_{011}^{011>11} = \frac{\frac{G}{2} + \frac{G-C}{2} + \frac{G-C}{2}}{3} = \frac{3G-2C}{6} = 0;$$

- *Fitness of phenotype* f_2. Individuals of phenotype f_2 fight 3 times, acting always like Hawks,

$$\Phi_{11}^{011>11} = \frac{G + \frac{G-C}{2} + \frac{G-C}{2}}{3} = \frac{2G-C}{3} = \frac{1}{4};$$

individuals of strategy 11 therefore are the ones that will increase, until they become the majority. In that situation, with a population with majority of $f_2 = 11$, i.e. "First Hawk, then Hawk" fitness result:

- *Fitness of phenotype* f_1. Individuals of phenotype f_1 fight a total of 3 times, the first time acting like Doves (0), the second and third acting like Hawks (11).

$$\Phi_{011}^{11>011} = \frac{0 + \frac{G-C}{2} + \frac{G-C}{2}}{3} = \frac{G-C}{3} = -\frac{1}{3};$$

- *Fitness of phenotype* f_2: Individuals of phenotype f_2 fight 2 times, acting always like Hawks,

$$\Phi_{11}^{11>011} = \frac{\frac{G-C}{2} + \frac{G-C}{2}}{2} = \frac{G-C}{2} = -\frac{1}{2}.$$

In a population with majority of strategist "First Hawk, then Hawk" the most cautious strategy "First Dove, then Hawk" is rewarded, bringing individual of type f_1 to increase.

The population will initially tend to oscillate and stabilize towards an equilibrium with the same number of individuals of both phenotypes (see Figure 7.3)

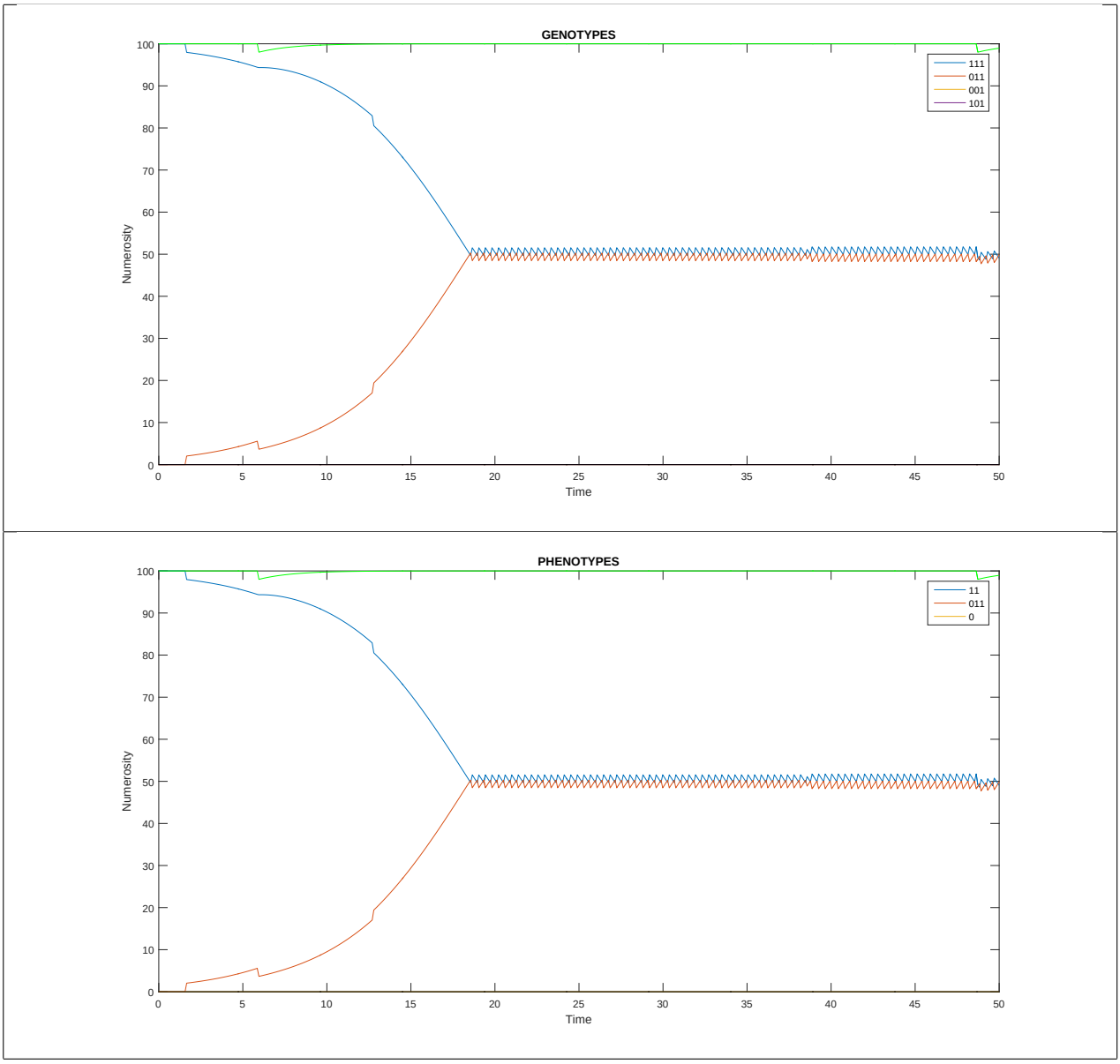


Figure 7.3: Genotypes and Phenotypes model (6.12)(6.13)(6.14) with Fibonacci GP-map. Time of execution is $T = 50$. The maximum intensity of the Poisson process is 0.1, there are no insertional mutations.

With the passage of time, however, the structure of the genome space $\mathcal{G}$ begins to determine a different equilibrium: any non-vital phenotype, although it may appear after a mutation, will not have time to mutate before disappearing, leaving room only for the neutral spaces of the vital phenotypes, distributed as shown in Figure 7.2. The neutral space of phenotype 11 has two elements, while 011 a single one, with 011 and 110 that can mutate only in 111. This situation leads to an imbalance that, for long times, leads to the dominance of the strategy 11, as can be seen in Figure 7.4.

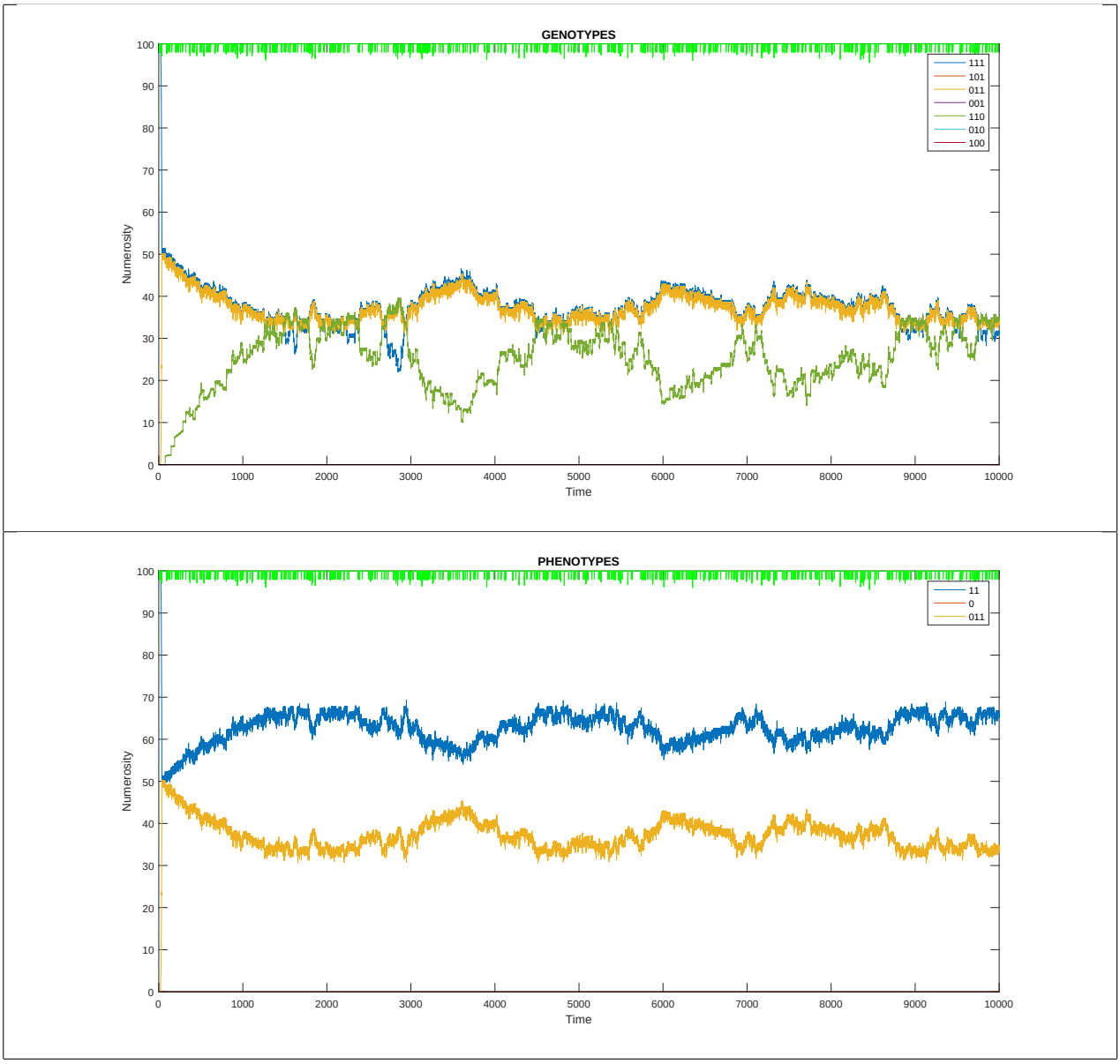


Figure 7.4: Genotypes and Phenotypes model (6.12)(6.13)(6.14) with Fibonacci GP-map. Time of execution is $T = 10000$. The maximum intensity of the Poisson process is 0.1, there are no insertional mutations.

When the probability of having an insertion mutation is positive, then new strategies can arise in the population, complicating the final result (Figure 7.5); the phenotype to affirm itself in the population is 11, with the largest neutral space. When the length of the genome becomes very large, the neutral space of the phenotype 11 grows exponentially in size (there are $2^{\tilde{L}-2}$ genomes of length $\tilde{L}$ that express this phenotype); since individuals in the same neutral space are indistinguishable from natural selection, after sufficient time each genome of this space will have about the same number of individuals. The Fibonacci GP-map therefore ceases to be representative when the number of genomes in the same neutral space is much

larger than the total number of individuals in the population, leading to the formation of populations with less than one individual for each genotype.

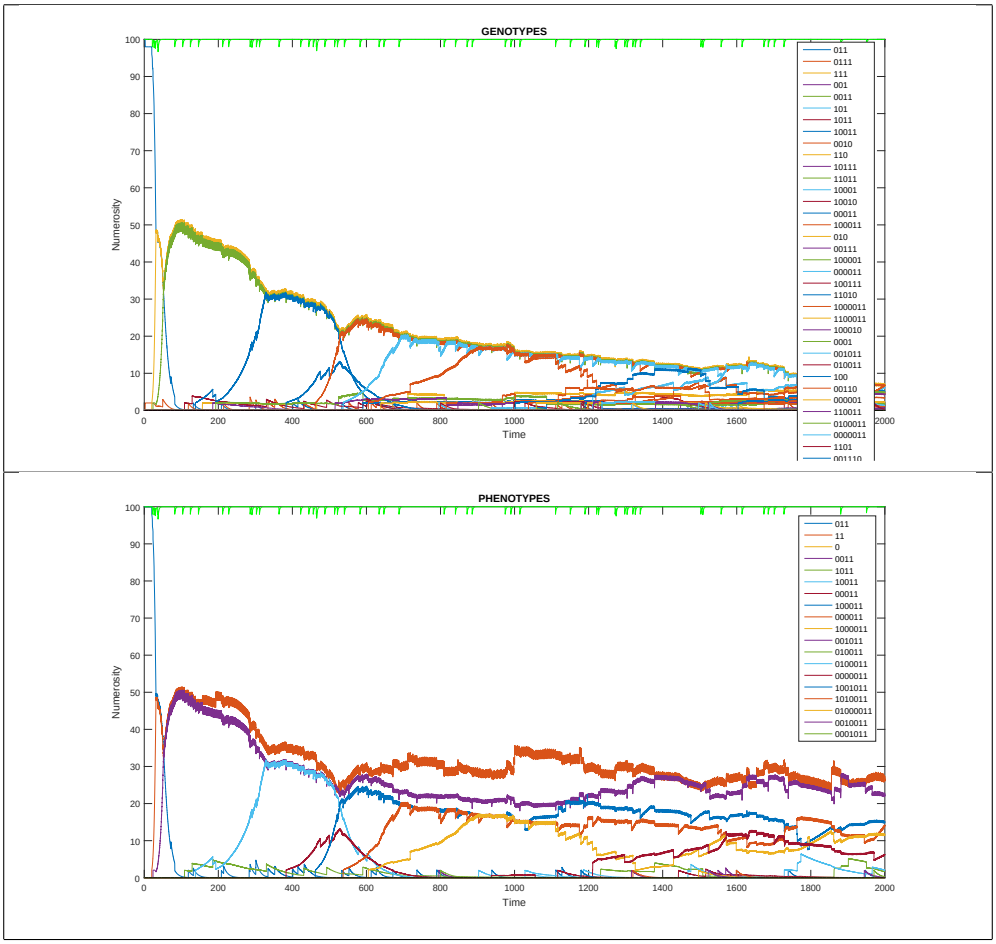


Figure 7.5: Genotypes and Phenotypes model (6.12)(6.13)(6.14) with Fibonacci GP-map. Time of execution is $T = 2000$. The maximum intensity of the Poisson process is 0.1, insertional mutations happen with probability $p = 0.1$.

Life-like cellular automaton

Here is an example showing the difference in action of mutation and selection in a clear way, using a Life-like cellular automaton. From [25]:

"One may define cellular automata on grids of high dimensions or on neighborhood structures more general than grids. The set of neighbors of a cell may be only those other cells nearest to it in the grid or may fall within a neighborhood of larger than unit radius. The state of a cell may depend only on the states of neighboring cells in the previous time step, or it may depend on the states of neighbors over several previous time steps. The number of states of each cell may be any finite number or even a continuously variable value, and researchers have considered update rules that are asynchronous, randomized, or quantum mechanical."

We restrict in this example our attention to set of rules that are very similar in structure to Conway's Game of Life [29]. Specifically, we consider binary cellular automata such that:

- The cells of the automaton form a two-dimensional square lattice.
- The neighbors of each cell are the eight lattice squares that are orthogonally or diagonally adjacent to it (Moore neighborhood).
- Each cell may be in one of two states, alive (1, black) or dead (0, white).
- All cells are updated simultaneously, and the time is discrete.
- In time step i, the state of any given cell is a function of the state of the same cell in time step $i-1$ and of the number of live neighbors it had in time step $i-1$.

In a cellular automaton of this type, a single cell may do one of four things within a single time step: if it was dead but becomes alive, then it is *born*; if it was alive and remains alive, then it *survives*; if it was alive and becomes dead, then it *dies*; and if it was dead and remains dead, then it is *quiescent*. The standard convention for naming these cellular automata is given by a sequence of characters in the form "$BxxxSyyy$". The xxx part of the rule string is a subset of the digits from 0 to 8, representing numbers of neighbors such that a dead cell with that many neighbors would become alive in the next time step, causing a birth event: the B stands for birth. The yyy part of the rule string is another subset of digits, representing numbers of neighbors such that a live cell with that many neighbors would remain alive in the next time step, causing a survival event: the S stands for survival.

For instance, Conway's Game of Life itself is represented by the rule string $B3S23$: a dead cell with three live neighbors leads to a birth event, and a live cell with two or three live neighbors leads to a survival event. All other combinations of cell state and number of neighbors lead to death or quiescence and can be inferred from the birth and survival parts of the rule string.

Genome space and mutations. The genome space $\mathcal{G}$ is $\{0,1\}^{L^2}$, for $L \geq 1$, set of the $L \times L$ grids with elements in $\{0,1\}$; for what regards the distance between genomes given by mutations, each element in the matrix of the genome g has probability p to change, changing from 0 to 1, with $p = k/L^2$, so that there are k average changes every mutation. So for all $g_1, g_2 \in \mathcal{G}$ $\mathbb{P}(g_1 \longmapsto g_2) > 0$ and

$$\mathbb{P}(g_1 \longmapsto g_2) = p^{d_H(g_1,g_2)} (1-p)^{L^2 - d_H(g_1,g_2)},$$

where d_H is the Hamming distance, calculated on the elements in $\mathcal{G}$.

GP-map and fitness. Starting from a genome g, the GP map consists into generate a cellular automaton that has g as initial state, for a fixed number of iterations: this automaton is the variant $B5678S45678$ of Conway's Game of Life: at discrete time steps, the status of each cell is updated based on the previous state of it and of the 8 cells that surround it; if the target cell is alive, it continues to be alive in the next iteration if it has 4 or more living cells surrounding it; if instead it is dead, it borns if it has 5 or more live cells sorrounding him. This particular rule brings the initial state to group together creating characteristic patterns, distinct zones of living cells of various sizes; survival is assured only for a high number of neighbors and birth only for 5 or greater, bringing solitary or slightly aggregated cells to disappear. This model is able to show well the different action of selection and mutation,

since the genome is a chaotic set of points, while the configurations generated by this variant of the game are similar to liveries of animals. Let define N_{max} the number of iterations chosen, and $\mathcal{H}$ the map that performs an iteration of the algorithm, then the GP-map results in:

$$GP := \mathcal{H}^{N_{max}} : \mathcal{G} \longmapsto \mathcal{F} \subset \mathcal{G}$$

$$g \in \mathcal{G} \longmapsto \mathcal{H}(g) \in \mathcal{G} \longmapsto \ldots \longmapsto \mathcal{H}^n(g) \in \mathcal{G} \longmapsto \ldots$$

Because the rule does not include births for 0, 1, 2 or 3 neighbors, then it is not possible for an initial pattern entirely contained in a bounding box to get out of it, so the algorithm eventually leads to periodic configurations [25]. From numerical simulations, choosing $L \leq 30$ and $N_{max} \geq 30$, any computed random initial data g_k converges to a fixed point $f_k = \mathcal{H}(f_k)$ in less than N_{max} iterations. From this consideration it is therefore possible to check how the GP-map generates a phenotype space $\mathcal{F} \subset \mathcal{G}$ of cardinality lower than $\mathcal{G}$. Infact, chosen an initial genome g_k which converges to f_k in less than N_{max} iterations, each element $\mathcal{H}^n(g_k)$ will express the same phenotype of g_k.

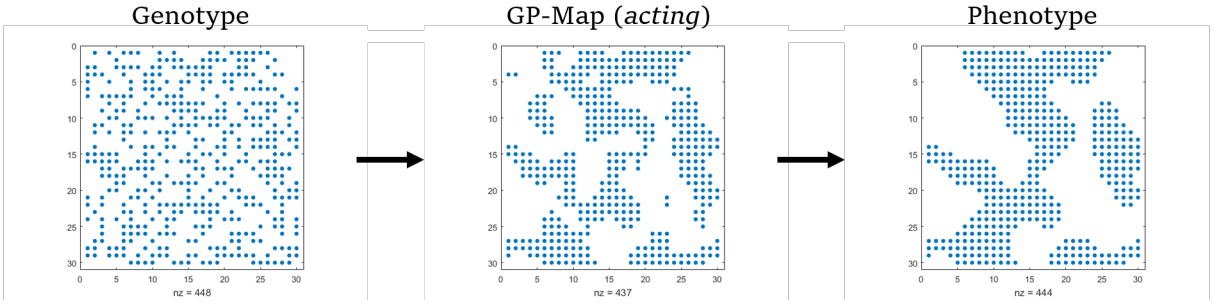


Figure 7.6: GP-map of Life like automaton model. The genotype, on the left, is a $L \times L$ matrix ($L = 30$) with elements in $\{0, 1\}$. The genotype is the inital status for the $B5678S45678$ version of Game of Life. The phenotype, on the right, is the equilibrium configuration (obtained in less than 30 iterations). The figure in the center is the cellular automaton in an iteration preceding the final one. nz counts the living cells of the square.

It is possible to define the fitness by working on the "apparent" characteristics of the phenotypes, evaluating the number of connected components and total area occupied: a binary array can be represented by a black and white image, where 1 represents a black pixel, and 0 a white pixel: two black pixels belong to the same connected component if it is possible to trace a path, moving between neighboring cells, which connects the two pixels; the number is then divided by the greatest number of connected components of an individuals in the population. The measure of the total area occupied is a measure of density, obtained by dividing the number of pixels equal to 1 for the total number of elements in the genome, L^2.

$$\Phi_k := \frac{\#\text{connected components of } f_k}{\max\{\#\text{connected components of } f, \text{ with } f \text{ in the population}\}} + \frac{\#\text{number of pixels of } f_k \text{ equal to } 1}{L^2}. \tag{7.1}$$

Examples. Even if it is finite, for $L = 20$ the cardinality of $\mathcal{G}$ is $2^{L^2} \sim 10^{120}$ elements. With the simulation times used, it was therefore possible to explore only a small subset of this space. In Figure 7.7 is possible to display the results of a simulation for $T = 1000$, obtained from an initial random genome. More simulations have been carried out, with different L, starting with the same initial genome (Figure 7.9 to 7.17), each time obtaining different phenotypes; albeit different from each other, the phenotypes obtained show all the characteristics favored by the selective pressure applied on them: greater number of connected components compared to the ancestral and higher density of living cells.

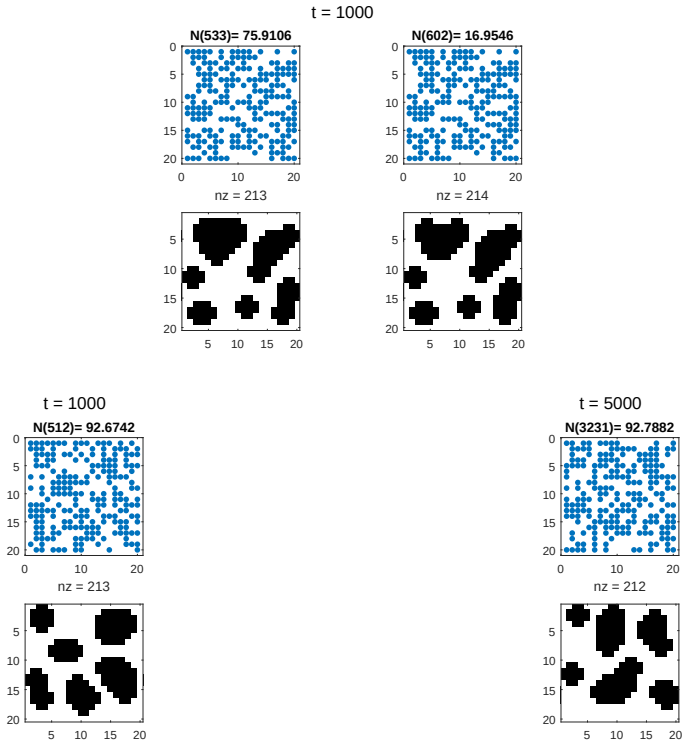
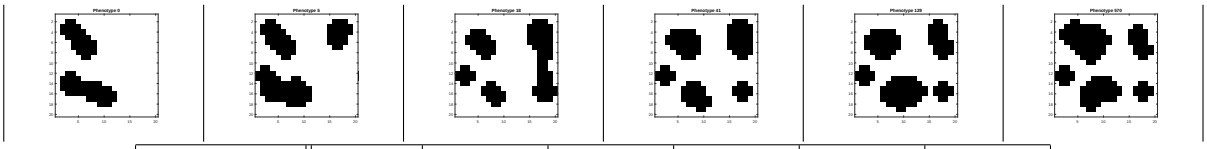
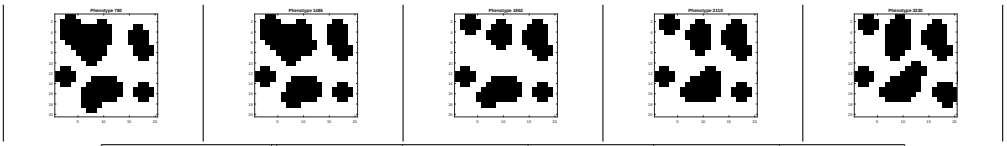


Figure 7.7: Genotypes and Phenotypes model (6.12)(6.13)(6.14) with Life-like automaton. The dimension of the grid is set at $L = 20$, the final time is $T = 1000$ for the first two run, $T = 5000$ for the third. We show the final phenotypes and genotypes of three different runs of the process starting from the same genome.

For the simulation in Figure 7.8, we have chosen as final time $T = 5000$ and we display the best phenotypes over time, reporting their fitness (increasing over time) and the time they appear the first time.



Phenotype	0	5	18	41	129	570
T	0	13.1	25.9	50.0	219.4	323.2
(cc, δ)	$(2, 75)$	$(3, 111)$	$(4, 113)$	$(5, 117)$	$(5, 133)$	$(5, 143)$
Φ	0.521	0.777	0.949	1.126	1.166	1.191



Phenotype	780	1485	1952	2110	3230
T	867.6	1850.9	2069.9	2489.9	4126.7
(cc, δ)	$(5, 155)$	$(5, 159)$	$(6, 130)$	$(6, 135)$	$(6, 156)$
Φ	1.221	1.231	1.325	1.337	1.390

Figure 7.8: Genotypes and Phenotypes model (6.12)(6.13)(6.14) with Life-like automaton. Simulation for $T = 5000$. The table shows the characteristics of all the phenotypes that, in a given time, have represented the majority of individuals; in the table, in the line *Phenotype* we indicate after how many mutations the represented phenotype has been generated; in the line T the first time after which it reached the majority in the population. In the line (cc, δ) we indicate respectively the number of connected components of the represented phenotype and the number of live cells of this configuration (with a maximum of 400). In the line Φ we indicate the fitness of the various phenotypes, imagining that they are in the same population (so in (7.1) $\max\{\#\text{connected components of } f, \text{ with } f \text{ in the population}\}$ is 6).

More simulations. In the following figures we report different executions of the simulation algorithm, showing the trend over time of the distribution of genotypes in the population, for $L = 10, 15, 20$.

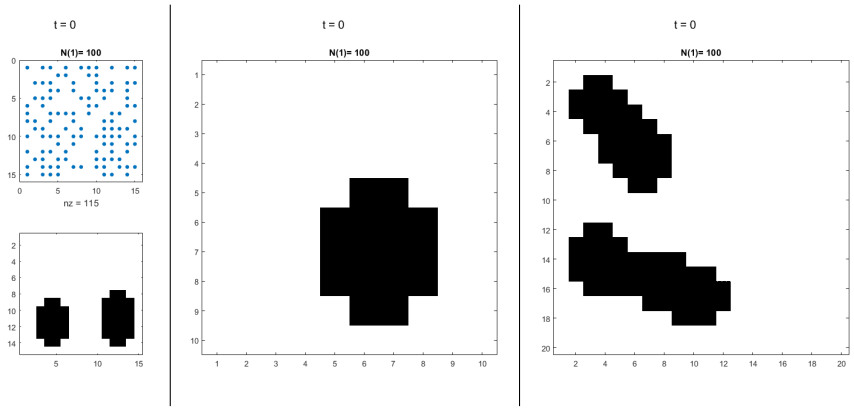
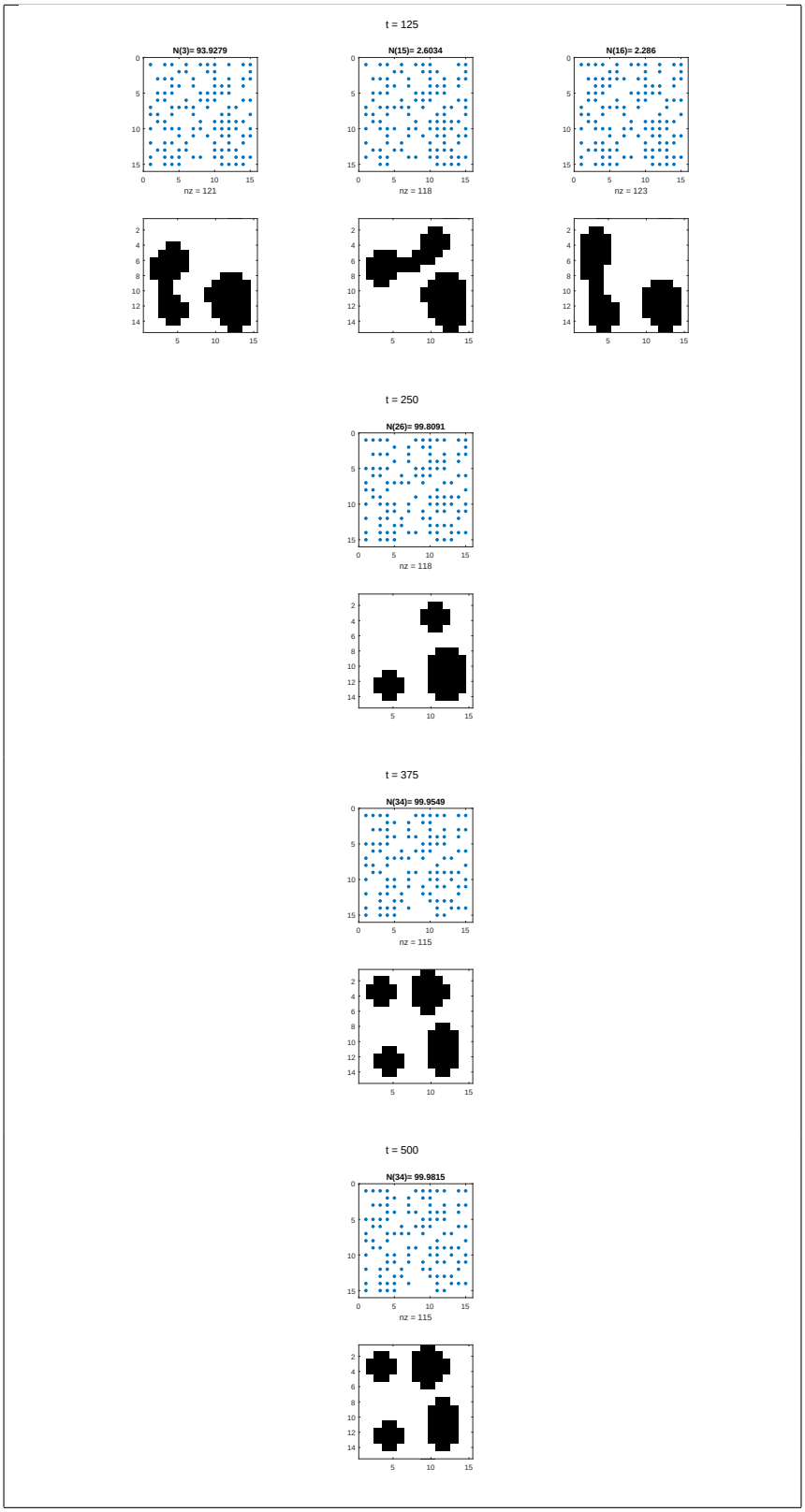


Figure 7.9: Initial phenotypes for subsequent simulations. In order, genotype and phenotype for $L = 15$, phenotype for $L = 10$, phenotype for $L = 20$.


t = 125
N(3)= 93.9279
N(15)= 2.6034
N(16)= 2.286
nz = 121
nz = 118
nz = 123
t = 250
N(26)= 99.8091
nz = 118
t = 375
N(34)= 99.9549
nz = 115
t = 500
N(34)= 99.9815
nz = 115

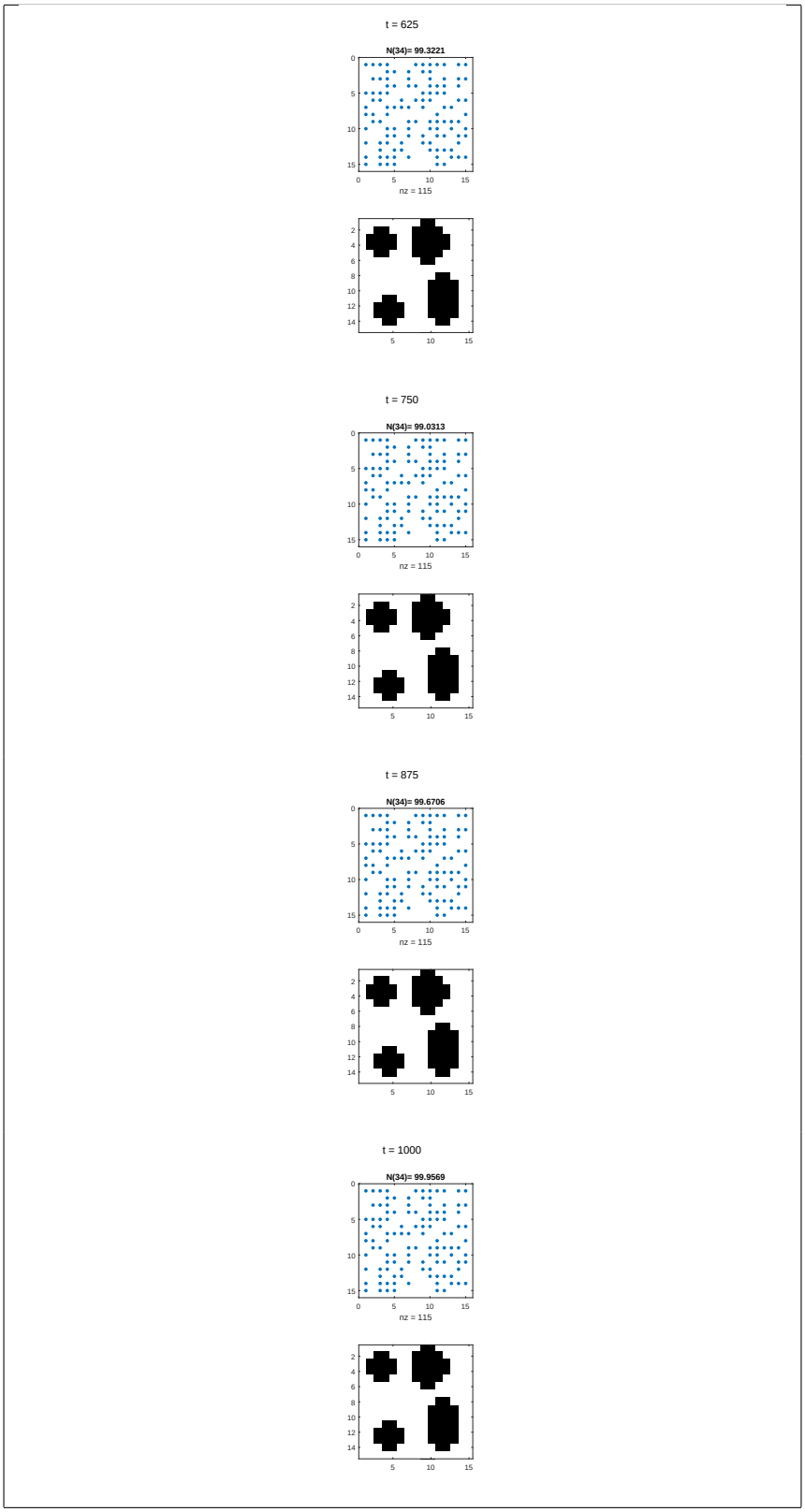
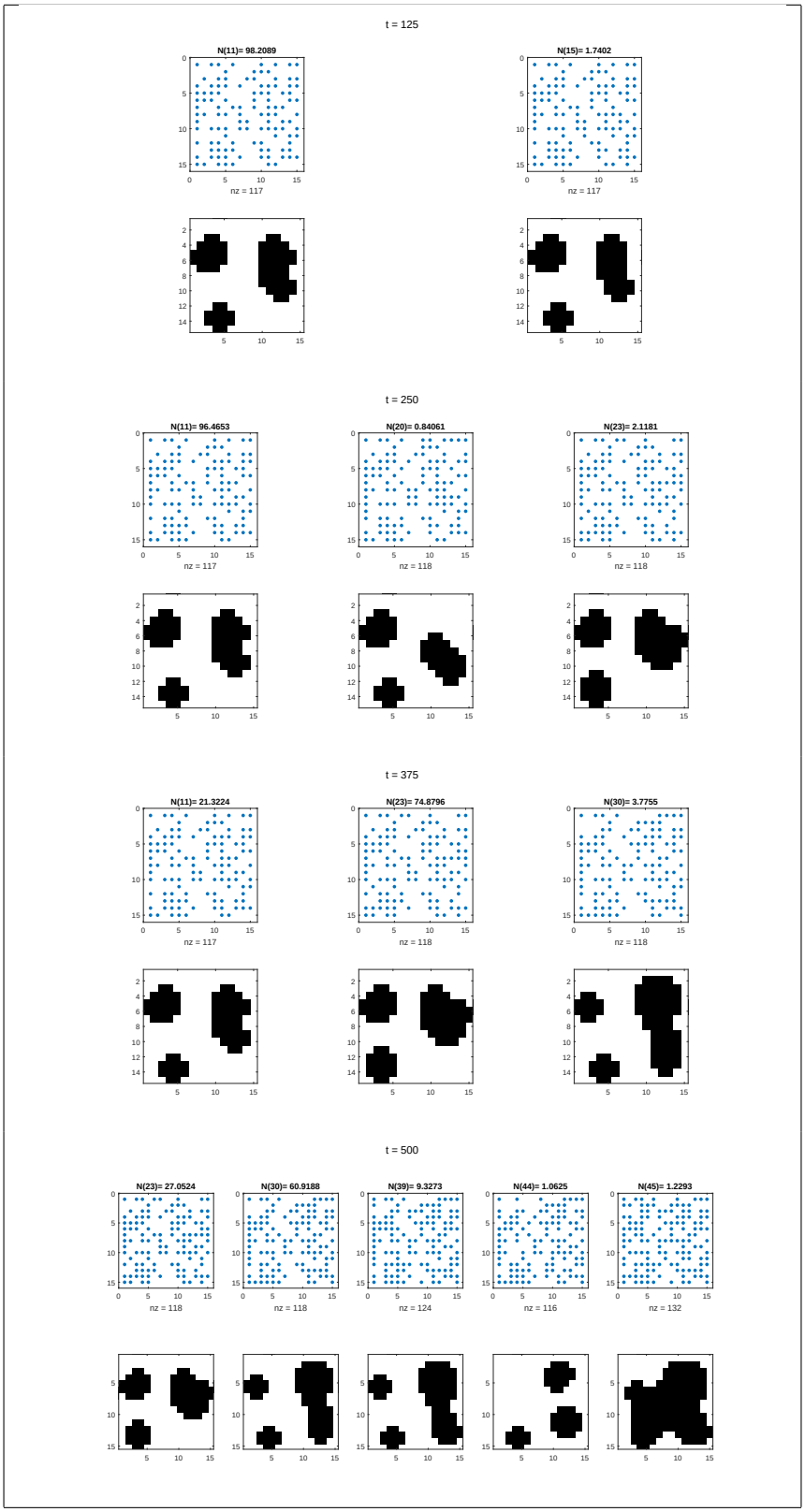


Figure 7.10: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. First simulation for $T = 1000$, dimension of the grid $L = 15$. The starting genotype and phenotype are shown in Figure 7.9. All genotypes and phenotypes with at least 1% of the maximum number of individuals are shown, with their number.


t = 125
N(11)= 98.2089
nz = 117
N(15)= 1.7402
nz = 117
t = 250
N(11)= 96.4653
nz = 117
N(20)= 0.84061
nz = 118
N(23)= 2.1181
nz = 118
t = 375
N(11)= 21.3224
nz = 117
N(23)= 74.8796
nz = 118
N(30)= 3.7755
nz = 118
t = 500
N(23)= 27.0524
nz = 118
N(30)= 60.9188
nz = 118
N(39)= 9.3273
nz = 124
N(44)= 1.0625
nz = 116
N(45)= 1.2293
nz = 132

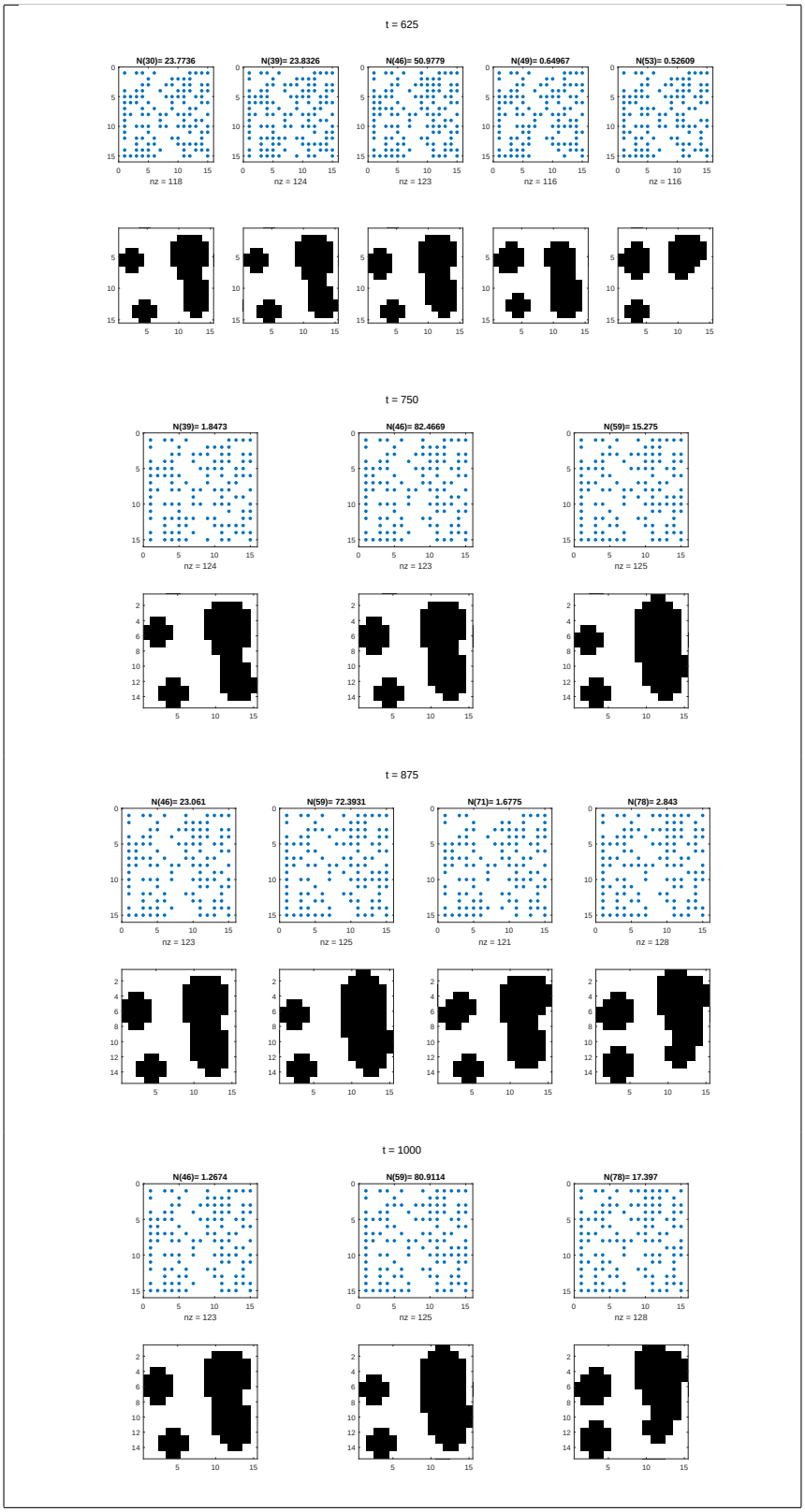
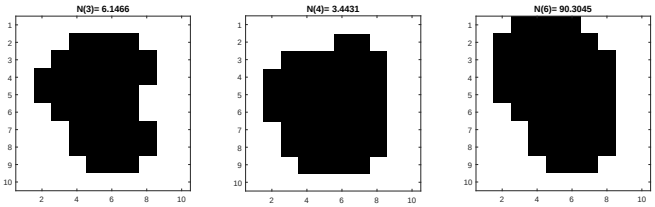
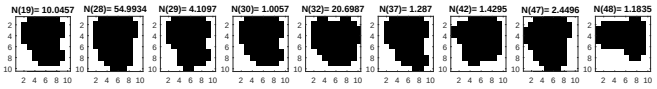


Figure 7.11: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. Second simulation for $T = 1000$, dimension of the grid $L = 15$. The starting genotype and phenotype are shown in Figure 7.9. All genotypes and phenotypes with at least 0.5% of the maximum number of individuals are shown, with their number.

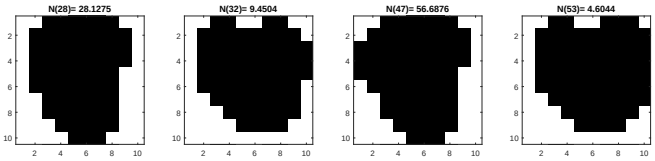
t = 125



t = 375



t = 500



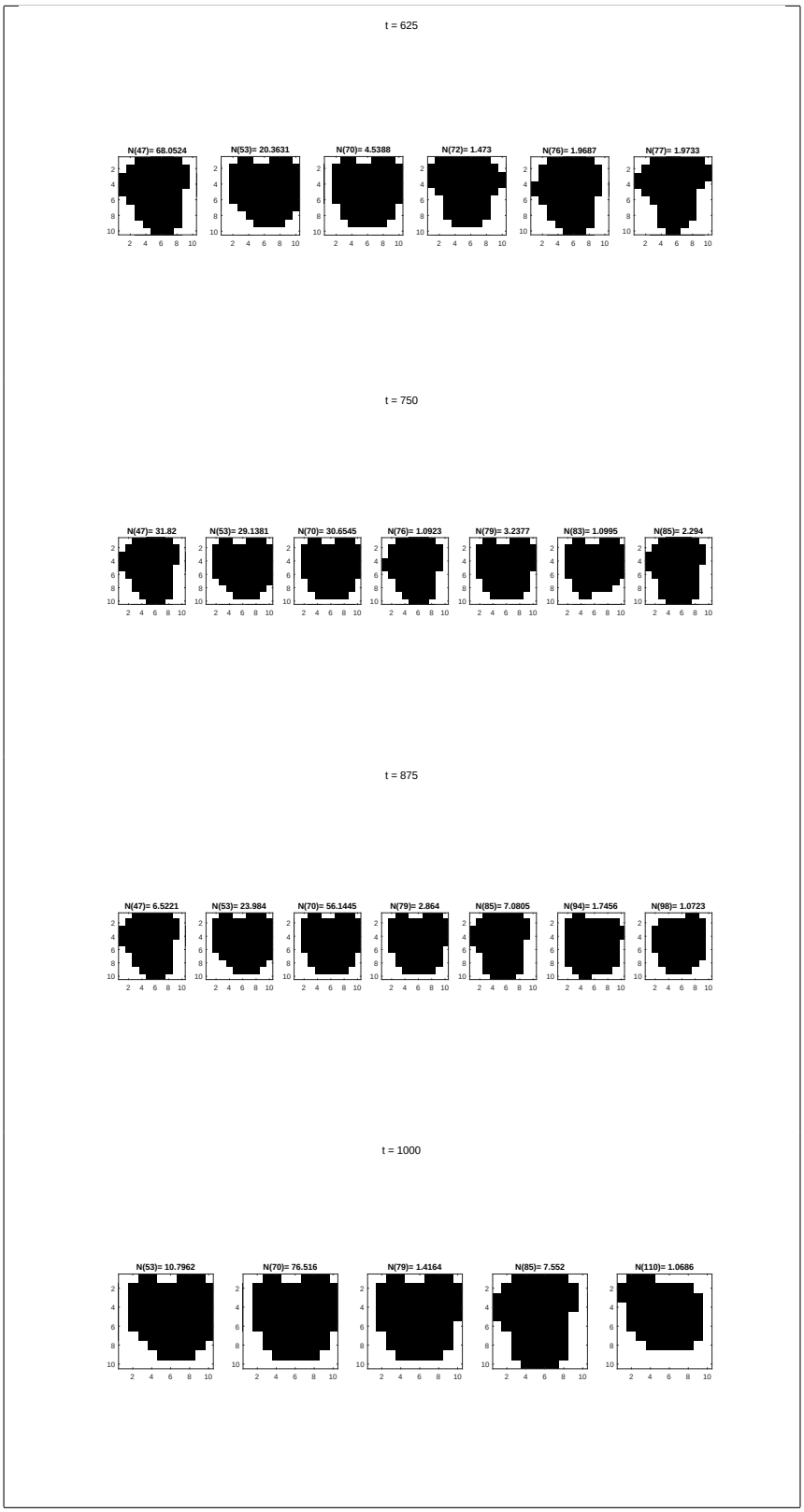
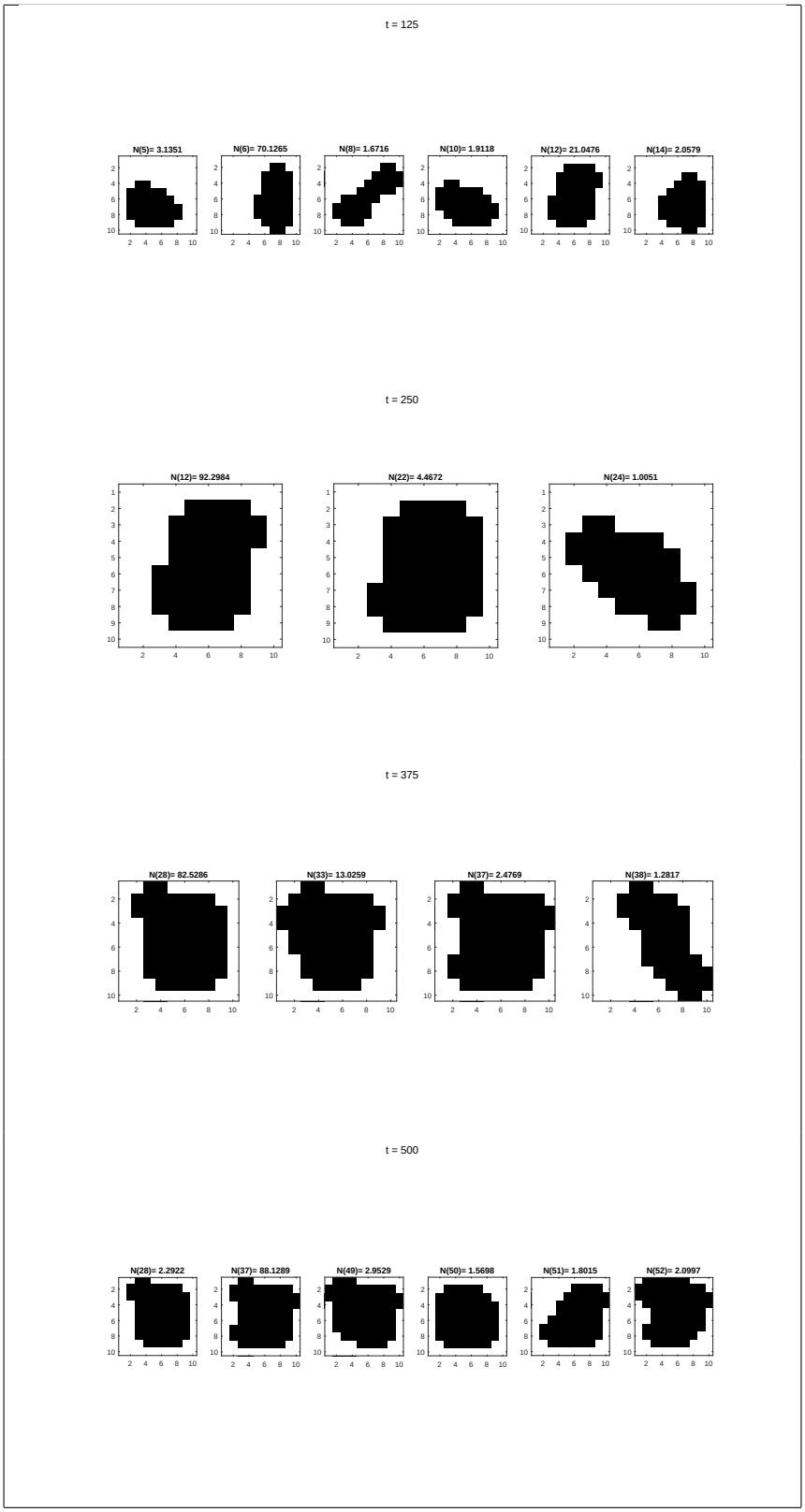


Figure 7.12: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. First simulation for $T = 1000$, dimension of the grid $L = 10$. The starting phenotype is shown in Figure 7.9. All phenotypes with at least 1% of the maximum number of individuals are shown, with their number.


t = 125
N(5)= 3.1351
N(6)= 70.1265
N(8)= 1.6716
N(10)= 1.9118
N(12)= 21.0476
N(14)= 2.0579
t = 250
N(12)= 92.2984
N(22)= 4.4672
N(24)= 1.0051
t = 375
N(28)= 82.5286
N(33)= 13.0259
N(37)= 2.4769
N(38)= 1.2817
t = 500
N(28)= 2.2922
N(37)= 88.1289
N(49)= 2.9529
N(50)= 1.5698
N(51)= 1.8015
N(52)= 2.0997

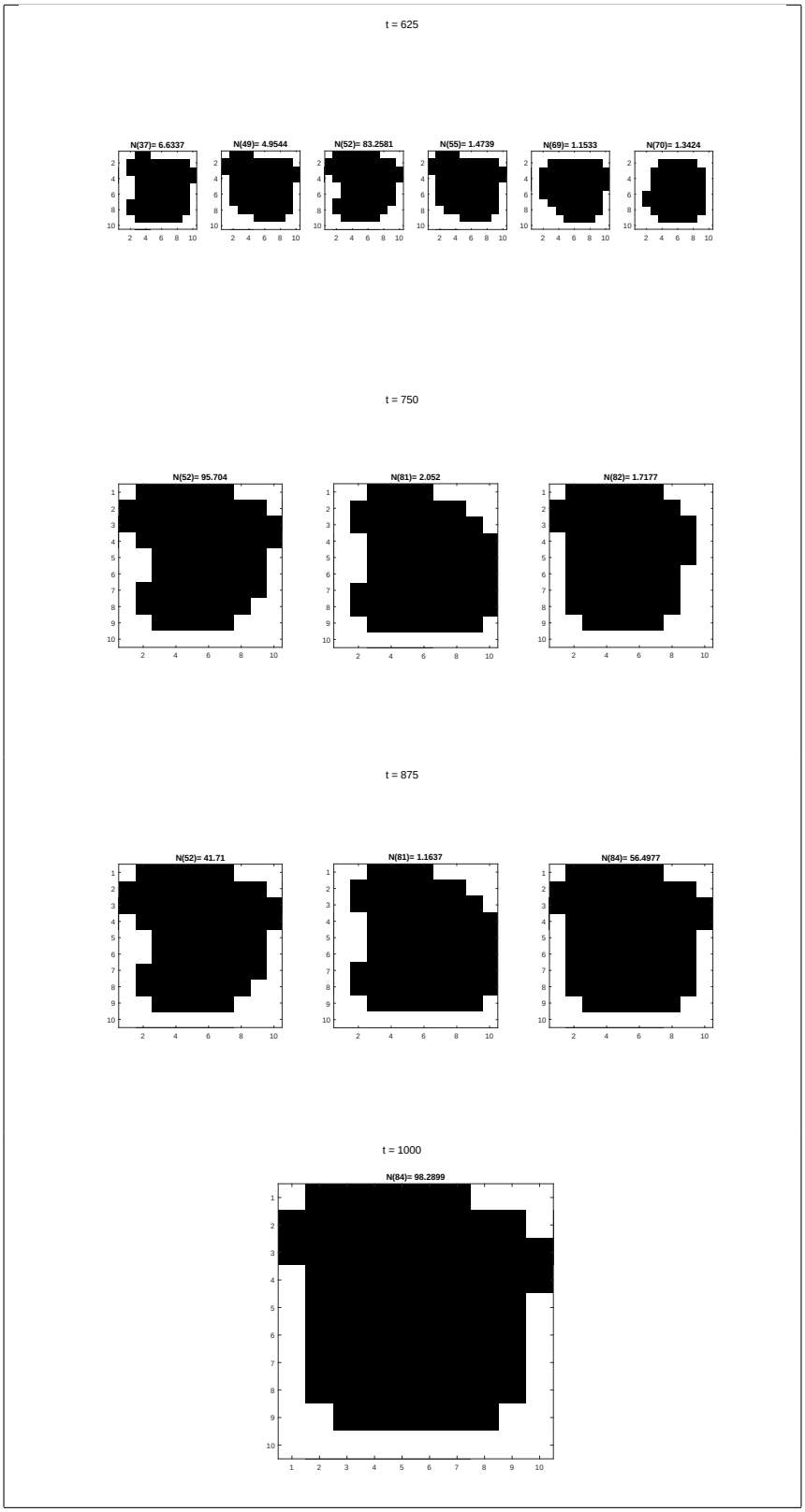
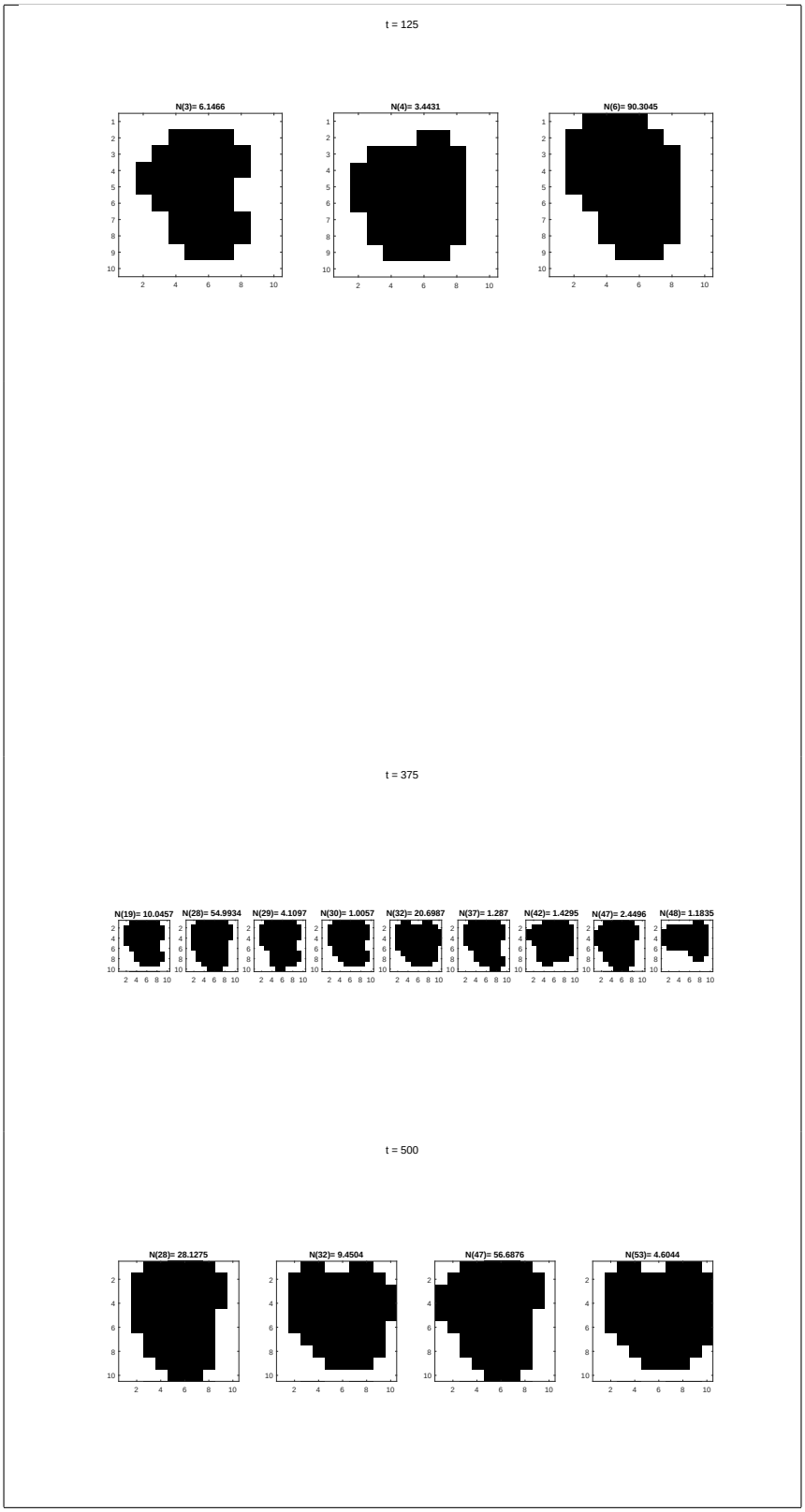


Figure 7.13: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. Second simulation for $T = 1000$, dimension of the grid $L = 10$. The starting phenotype is shown in Figure 7.9. All phenotypes with at least 1% of the maximum number of individuals are shown, with their number.


t = 125
N(3)= 6.1466
N(4)= 3.4431
N(6)= 90.3045
t = 375
N(19)= 10.0457
N(28)= 54.9934
N(29)= 4.1097
N(30)= 1.0057
N(32)= 20.6987
N(37)= 1.287
N(42)= 1.4295
N(47)= 2.4496
N(48)= 1.1835
t = 500
N(28)= 28.1275
N(32)= 9.4504
N(47)= 56.6876
N(53)= 4.6044

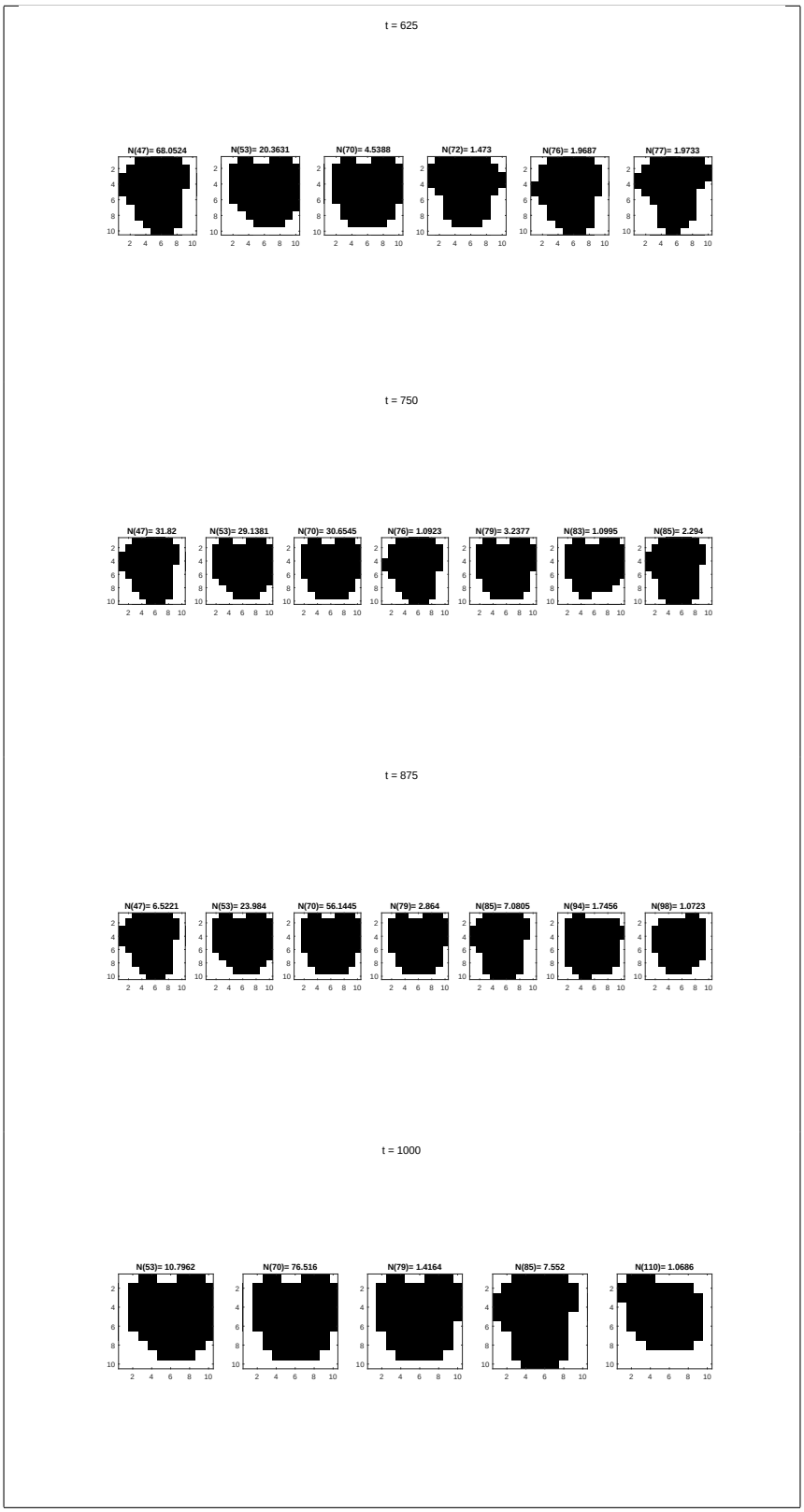
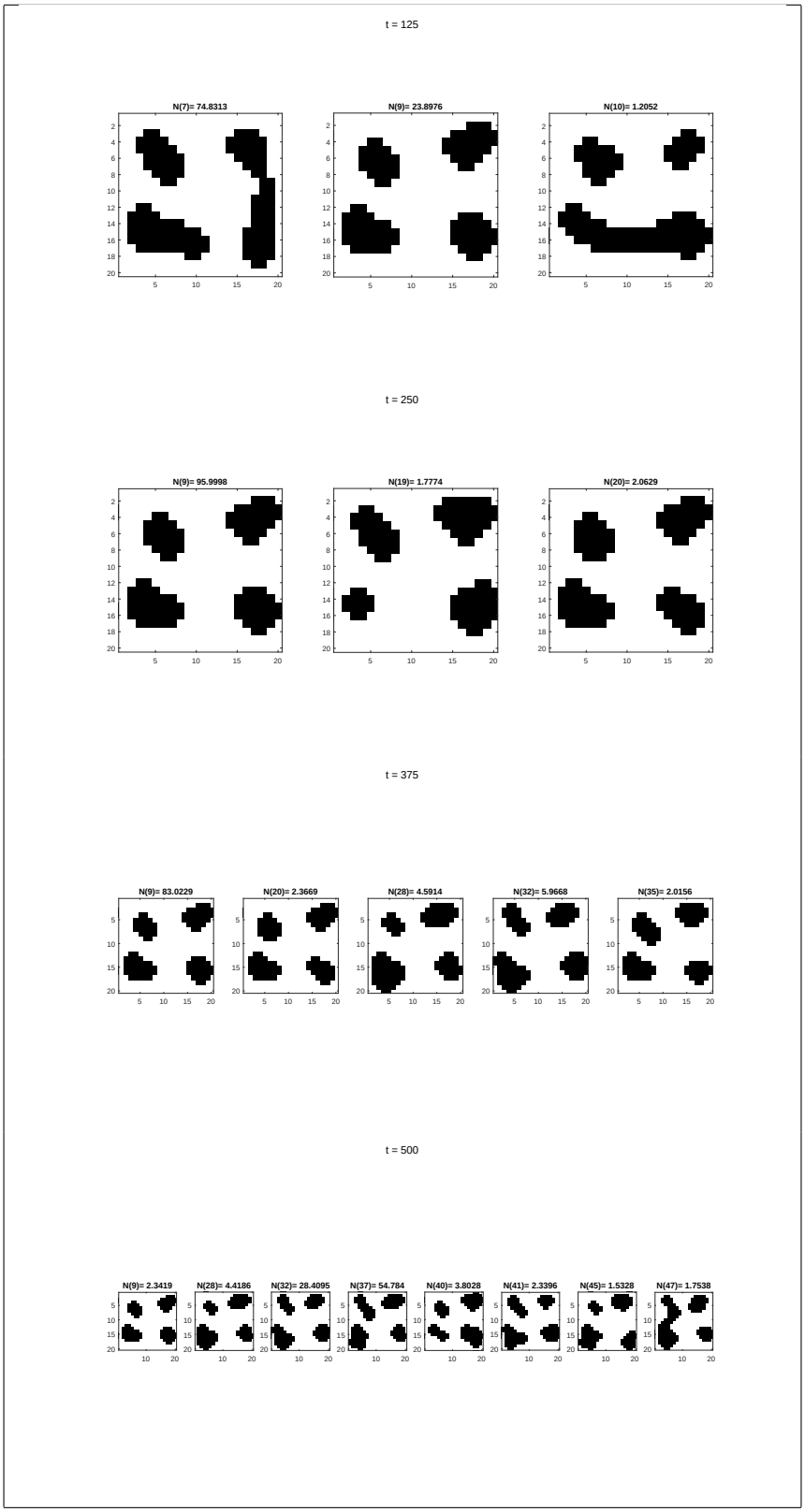


Figure 7.14: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. Third simulation for $T = 1000$, dimension of the grid $L = 10$. The starting phenotype is shown in Figure 7.9. All phenotypes with at least 1% of the maximum number of individuals are shown, with their number.


t = 125
N(7)= 74.8313
N(9)= 23.8976
N(10)= 1.2052
t = 250
N(9)= 95.9998
N(19)= 1.7774
N(20)= 2.0629
t = 375
N(9)= 83.0229
N(20)= 2.3669
N(28)= 4.5914
N(32)= 5.9668
N(35)= 2.0156
t = 500
N(9)= 2.3419
N(28)= 4.4186
N(32)= 28.4095
N(37)= 54.784
N(40)= 3.8028
N(41)= 2.3396
N(45)= 1.5328
N(47)= 1.7538

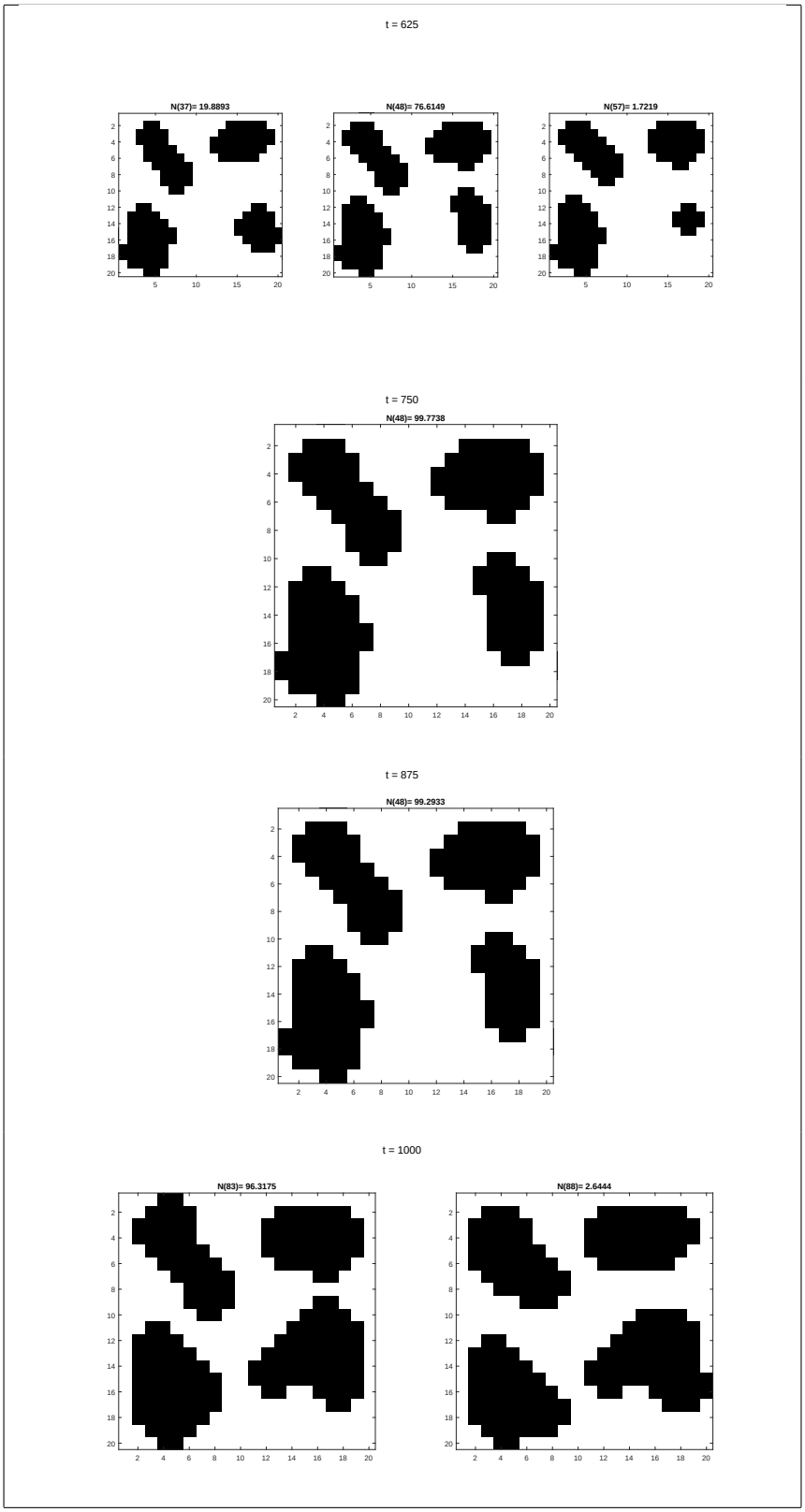
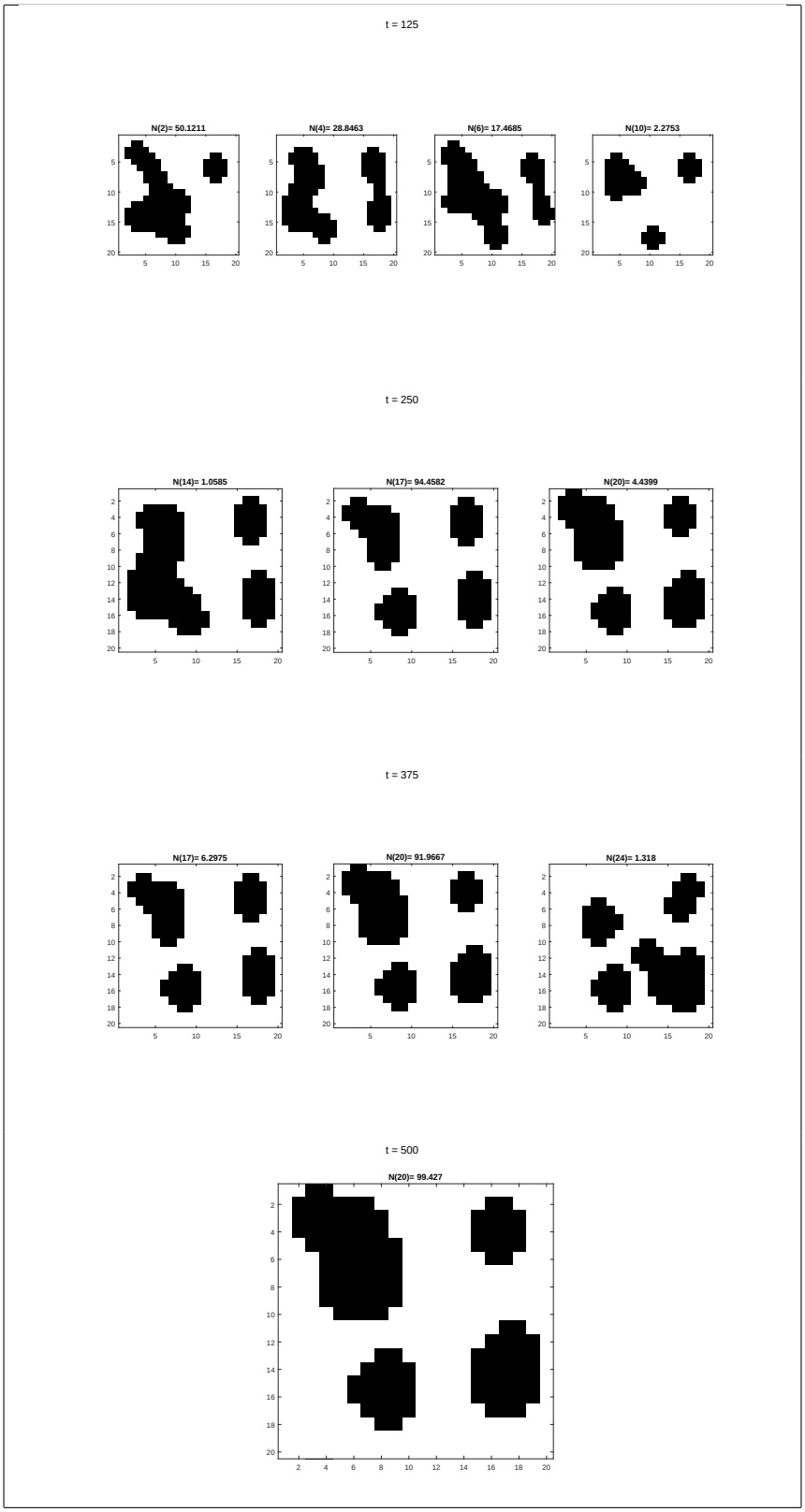


Figure 7.15: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. First simulation for $T = 1000$, dimension of the grid $L = 20$. The starting phenotype is shown in Figure 7.9. All phenotypes with at least 1% of the maximum number of individuals are shown, with their number.


t = 125
N(2)= 50.1211
N(4)= 28.8463
N(6)= 17.4685
N(10)= 2.2753
t = 250
N(14)= 1.0585
N(17)= 94.4582
N(20)= 4.4399
t = 375
N(17)= 6.2975
N(20)= 91.9667
N(24)= 1.318
t = 500
N(20)= 99.427

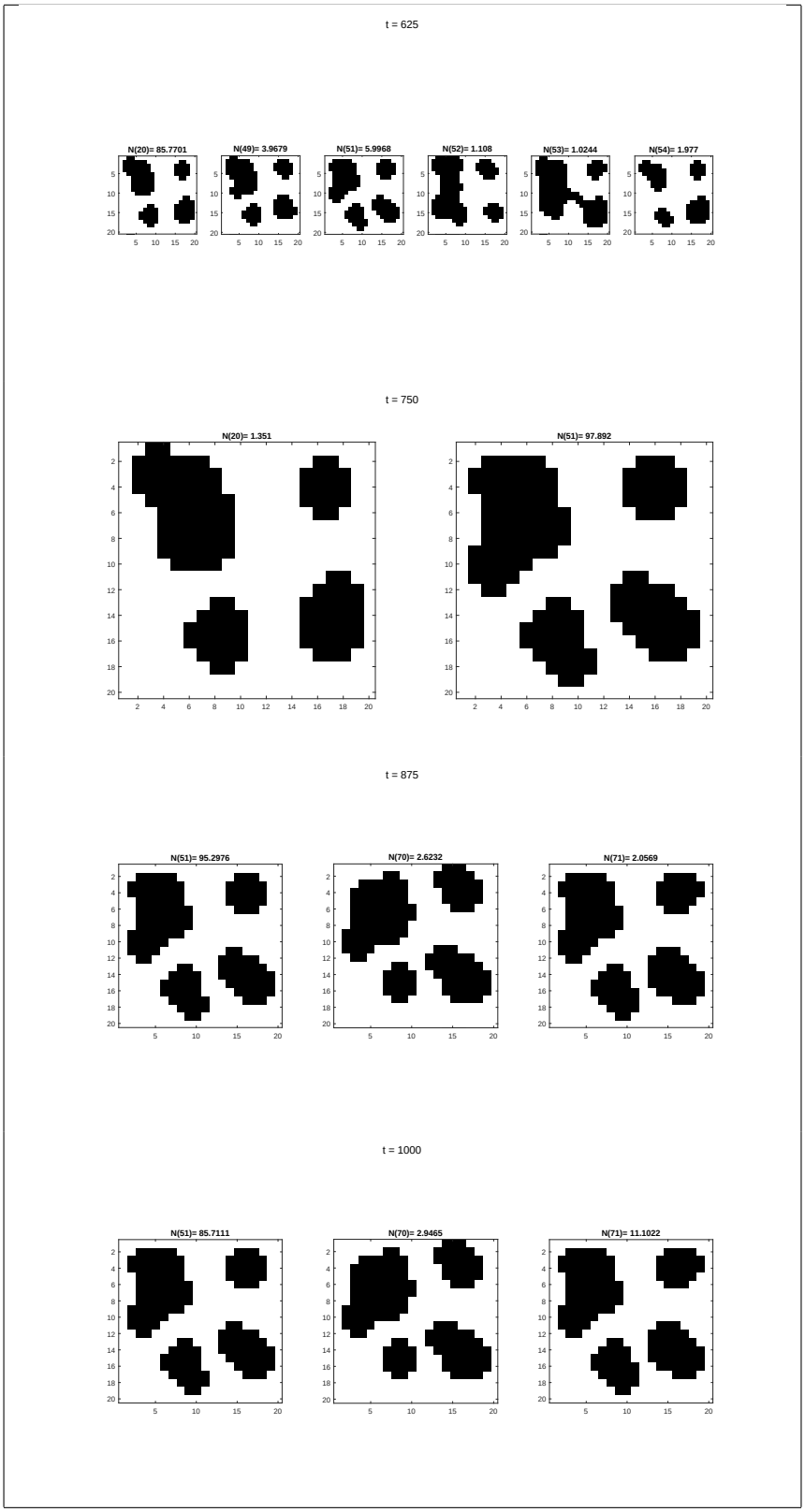
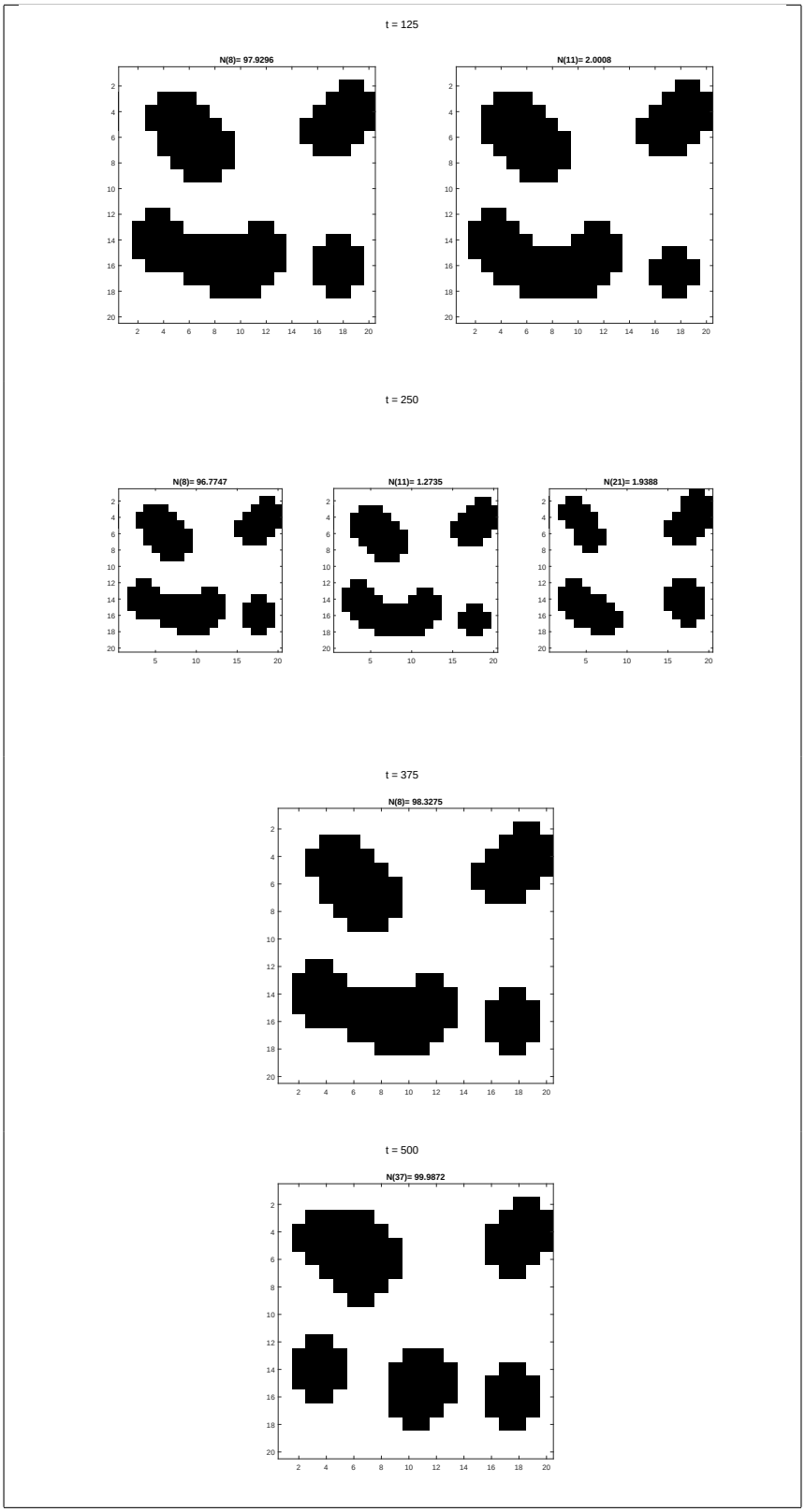


Figure 7.16: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. Second simulation for $T = 1000$, dimension of the grid $L = 20$. The starting phenotype is shown in Figure 7.9. All phenotypes with at least 1% of the maximum number of individuals are shown.


t = 125
N(8)= 97.9296
N(11)= 2.0008
t = 250
N(8)= 96.7747
N(11)= 1.2735
N(21)= 1.9388
t = 375
N(8)= 98.3275
t = 500
N(37)= 99.9872

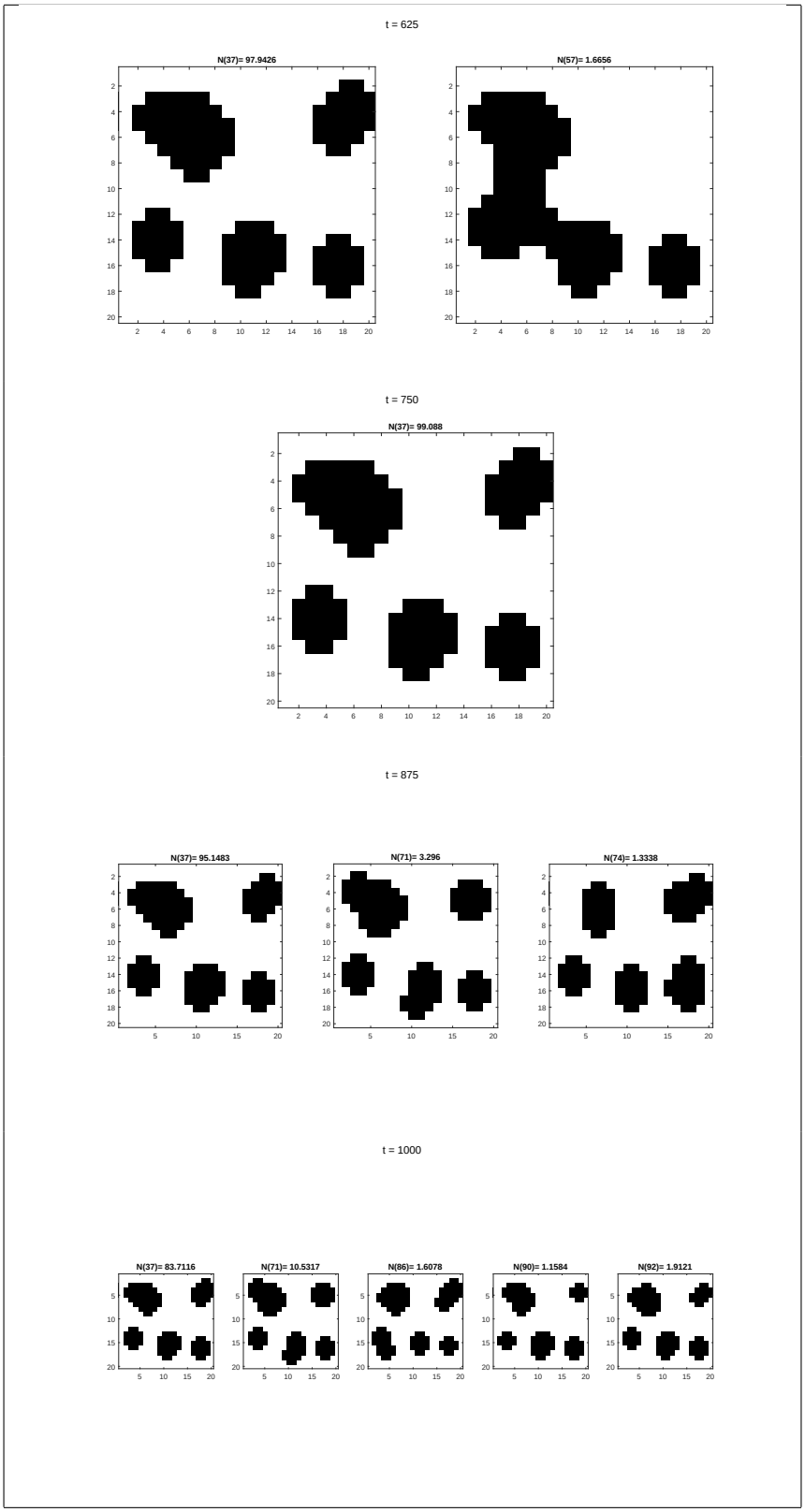


Figure 7.17: Phenotypes for model (6.12)(6.13)(6.14) with Life-like automaton. Third simulation for $T = 1000$, dimension of the grid $L = 20$. The starting phenotype is shown in Figure 7.9. All phenotypes with at least 1% of the maximum number of individuals are shown.

Replicator mutator

In this example, we consider a finite genome space, composed by only 4 elements. We show numerically how convergence at equilibrium is slower if mutations occur at a different level of selection.

Genome space and mutations. The genome space in this simplified model is $\mathcal{G} = \{1, 2, 3, 4\}$, there are only 4 possible genomes in the population; one genome can mutate into another if it has exactly 1 different base, i.e. according to the jump matrix Π of a random walk on $\mathcal{G}$:

$$\mathcal{G} = \{1, 2, 3, 4\} \tag{7.2}$$

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

GP-map and fitness. The GP map generates two different phenotypes, A and B, with $GP(i) = A$ for $i = 1, 2, 3$ and $GP(4) = B$; on these phenotypes we consider constant fitness, $f_A = 1$ and $f_B > f_A$ (see Figure 7.18).

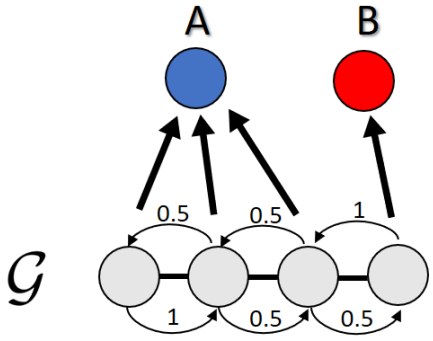


Figure 7.18: Graph of the Markov chain on the genome space $\mathcal{G}$, given by the matrix Π; the GP map with the neutral spaces are displayed, with phenotype A in blue, B in red.

Examples. In this model, all individuals are equally likely to mutate, as the temporal frequency of mutations of a genotype g_i is proportional only to its frequency in the population. We decide to compare this model with the simplest example of quasispecies (see Chapter 4), (4.1):

$$\begin{cases} \dot{x}_k = \left(\Phi_k - \bar{\Phi}\right) x_k + \sum_{i=1}^{n} \Phi_i m_{ik} x_i, \\ k = 1, \ldots, n, \end{cases}$$

with 2 types, imposing the probability that replication of individuals of a type results in individuals of the other equal, i.e. the mutation terms are the same, $m_0 := m_{12} = m_{21}$, obtaining the differential equation (reducing the variables to x_1 only):

$$\dot{x}_1 = (f_1 - f_2)\, x_1(1 - x_1) + f_2 m_0\, (1 - x_1) - f_1 m_0 x_1.$$

We compare the above equation with a version of model (7.2) that has no structure in the genotype space. So let $\mathcal{G}_1$ be the space of genomes $\{1, 2\}$, each of which can mutate into the other. If we take the identity as the GP-map, there is no difference between genotype i and the phenotype it expresses (which we can call i, see Figure 7.19), it can be assumed that fitness acts on the genotype.

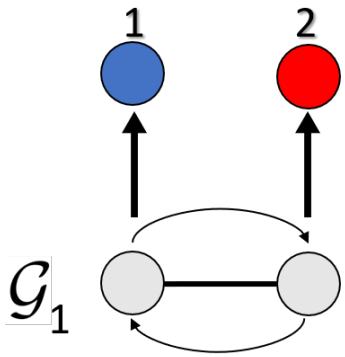


Figure 7.19: Graph of the simplest case of genome space $\mathcal{G}_1$, with two genotypes that have the same probability to mutate in each other. The GP-map is the identity.

In Figure 7.20 we simulate both models, obtaining a value for m_0 that produces similar transient and final equilibrium.

We then replace the space $\mathcal{G}_1$ with $\mathcal{G}$ and imagine two different situations: in both initially there are only individuals that show the phenotype A, with lower fitness; in the first case, the individuals are of genotype 3, which has probability 1/2 to mutate into genotype 4; in the second case instead, the individuals are of genotype 1, and can generate individuals of phenotype B with lower probability, mutating first in 2, then in 3, finally in 4. Although they are equivalent (phenotypically), the presence of a *structure* in the genotype space changes the trend of the process. In both cases, eventually, the genotype 4 will be generated and, having higher fitness, will grow in number, but in the second case we expect that individuals at lower fitness will last for longer. In Figure 7.21 the result is shown when the initial genome is chosen randomly, while in Figure 7.22 the initial genomes are chosen, one at a time, 4, 3 and 2. The convergence time at equilibrium can be then modified by the structure of the space of the genomes.

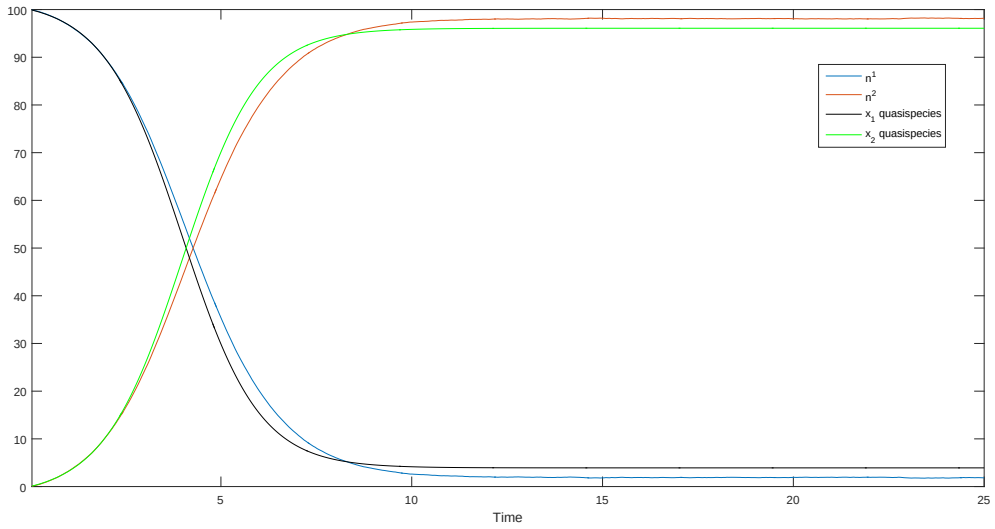


Figure 7.20: Quasispecies model for 2 types, in green and red, compared to expected value of genotype-phenotype model (averaged on 500 independent runs) in red and blue. $\lambda(N) = 1$, therefore the process underlying the mutations is homogeneous. The mutation term of quasispecies is $m_0 = 0.02$. Fitness $\Phi_1 = 1$, $\Phi_2 = 2$, initial population is formed only by individuals of type 1. $T = 25$.

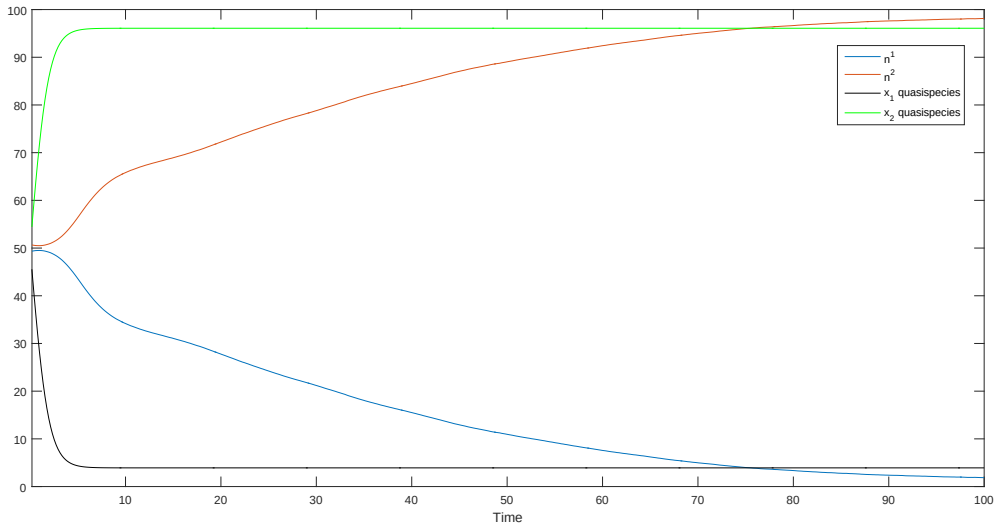


Figure 7.21: Quasispecies model for 2 types, in green and red, compared to expected value of genotype-phenotype model (averaged on 5000 independent runs) in red and blue. $\lambda(N) = 1$, therefore the process underlying the mutations is homogeneous. The mutation term of quasispecies is $m_0 = 0.02$. Fitness $\Phi_1 = 1$, $\Phi_2 = 2$. The initial density of quasispecies is equally distributed between the two initial types; during each run of the genotype-phenotype model, the initial population is formed by individuals of the same genotype i, $i \in \{1, 2, 3, 4\}$, with i chosen randomly, so as to have, statistically, uniform distribution of initial *phenotypes*: so the probability of having as initial genotype 4 (the only one expressing the phenotype 1) is three times greater than having one of the others as the initial genotype. $T = 100$.

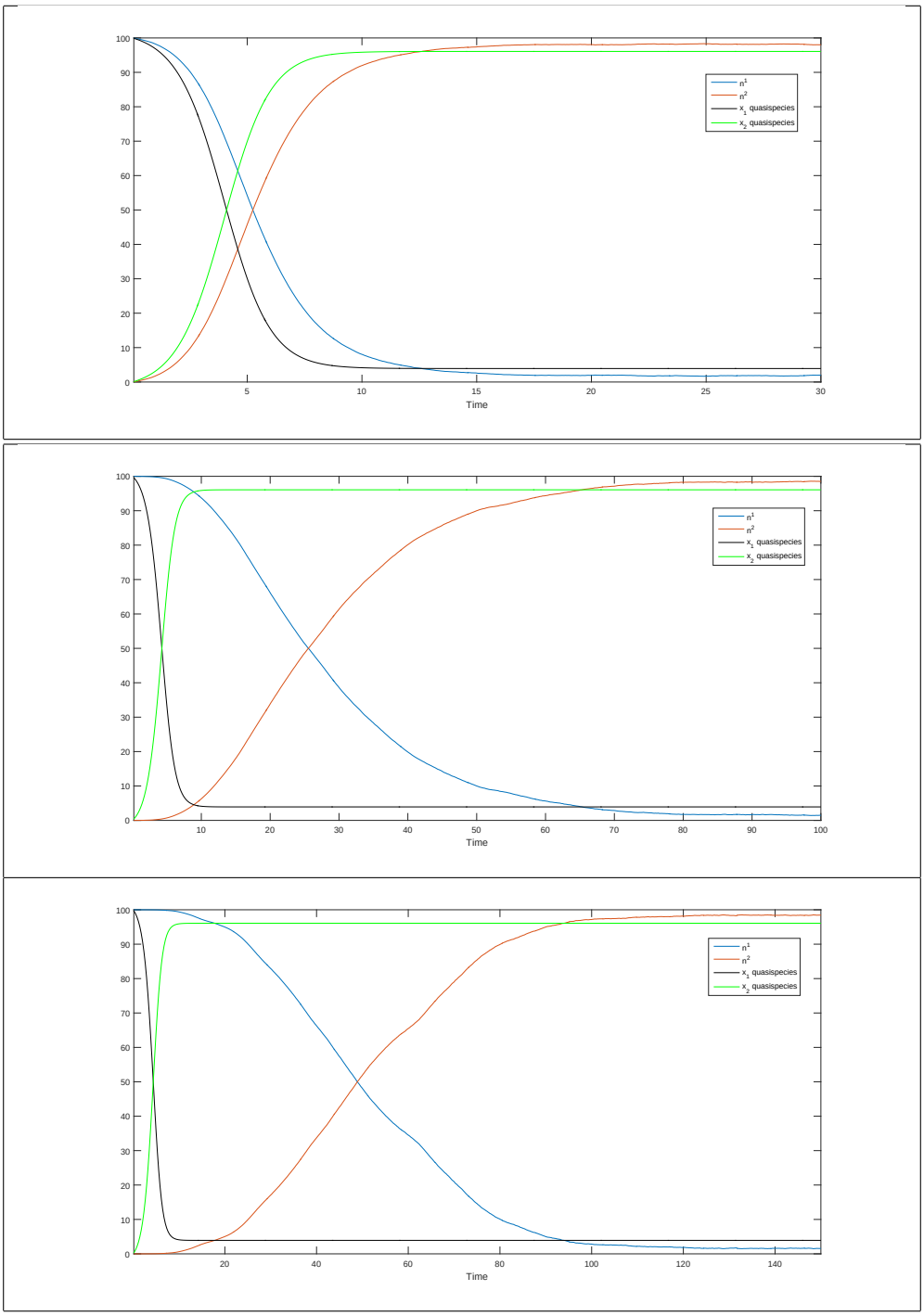


Figure 7.22: Quasispecies model for 2 types, in green and red, compared to expected value of genotype-phenotype model (averaged on 500 independent runs) in red and blue. $\lambda(N) = 1$, therefore the process underlying the mutations is homogeneous. The mutation term of quasispecies is $m_0 = 0.02$. Fitness $\Phi_1 = 1$, $\Phi_2 = 2$. The initial population of quasispecies is divided equally between the two initial types; the initial population of each run of the genotype-phenotype process is different in each figure. Respectively, in the first figure the initial population is made up exclusively of individuals of genotype 3, in the second of genotype 2, in the third of genotype 1. The more the initial genotype is "distant" from 4, which expresses a different phenotype, the longer it takes the system to converge to equilibrium (final times are, respectively, 30, 100, 150).

7.3 Conclusions

In this chapter we dealt with the problem of finding a model for the theory of Evolution, which starts from the concept of rare mutations shown in [4] and in Chapter 5. We have proposed a model for the dynamics of individuals of genotypes, defining a space of potential genomes $\mathcal{G}$, which can also have infinite cardinality. We have therefore constructed a mutation process that acts on the genotypes; such a process explores the space $\mathcal{G}$, producing over time individuals with features never seen before. Another step was to differentiate the effect of the selection, making it act on the phenotype, that is the environmental expression of a genotype. Two genotypically different individuals may express the same phenotypes, and for this reason subjected to same selective pressures from the environment. A phenotype is therefore characterized not only by a fitness, but represents a whole space of genotypes that produce it, which can have different size; the mere presence of such a space can therefore alter the dynamics of selection and adaptation, modifying the proportions of individuals or the times of convergence to equilibrium.

Further developments, in line with the model developed in Chapter 5, could concerns dynamics in which spatiality has a role. In this model of ordinary equations there is a spatially homogeneous population, which we can imagine concentrated in one point. To study a model in which various populations, arranged at different points in a geographic region, interact with each other through gene exchanges could lead to consider immigration and emigration phenomena, invasion of more fitted types. It would be possible to imagine different environmental conditions in different points of the space, obtaining different fitness landscapes. This idea could be developed by considering populations as nodes in a network, connected by flows of genetic informations, or as functions of space, thus developing models of partial derivatives equations.

We could then focus on the dynamics of extinction and low number of individuals. In this model, in fact, it is not possible for one type of individuals to become extinct: even when it is disadvantaged by selection, at most the decrease is exponential, leading to disappearance only in infinite time. The only case in which we can have an extinction is after a mutation, when a type, already very rare in the population, is transformed completely into a new one. When the number decreases, we could think of a model that turns into particle-based, in which individuals are treated as single entity, so that we can take into account phenomena such as bottlenecks, founder effect and genetic drifts, related to statistical fluctuations of populations composed of few individuals [35, 61].

Both of these ideas deserve further investigations.

Part III

Meta-heuristic algorithms inspired by Nature

Chapter 8

Introduction and Algorithms

This part, composed of two chapters, is the extended version of a paper written with Giovanni Sebastiani (IAC-Rome) about a procedure (called *restart procedure*) for the optimization of opportune combinatorial optimization problems, which is currently accepted [83]. The basic algorithms of the work, that are *Ant Colony optimization* and *Genetic algorithms* are of explicit natural inspiration, and are presented in this Chapter. Then, in Chapter 9, the real problem is addressed; after the first part of the theoretical treatment of the problem, there is a Section where the nature-inspired algorithm are numerically implemented and used together with the new procedure. In general, solving a *combinatorial optimization problem* (COP) consists of finding an element, within a finite search domain, which minimizes a given *fitness function*. The domain has typically a combinatorial nature, e.g. the space of the hamiltonian paths on a complete graph. The COP prototype is the Traveling Salesman Problem (TSP), whose solution is a Hamiltonian cycle on a weighted graph with minimal total weight [6]. Although a solution of a COP always exists, finding it may involve a very high computational cost. The study of the computational cost of numerical algorithms started in the early 1940s with the first introduction of computers. Two different kinds of algorithms can be used to solve a COP problem: exact or heuristic. A method of the former type consists of a sequence of non-ambiguous and computable operations producing a COP solution in a finite time. Unfortunately, it is often not possible to use exact algorithms. This is the case for instances of a $\mathcal{NP}$-complete COP. In fact, to establish with certainty if any element of the search space is a solution, requires non-polynomial computational cost. Alternatively, heuristic algorithms can be applied. Such type of algorithms only guarantee either a solution in an infinite time or a suboptimal solution. Of great importance are the *meta-heuristic* algorithms (MHA) [9] which are algorithms that are independent of the particular COP considered, and often stochastic. Among them, there are Simulated Annealing [52], Tabu Search [33], Genetic Algorithms [34] and Ant Colony Optimization (ACO) [23].

8.1 Ant Colony Optimization and Genetic Algorithms

8.1.1 Ant Colony Optimization

Ant Colony Optimization (ACO) is a recently developed, population-based approach which has been successfully applied to several NP-hard combinatorial optimization problems. As the name suggests, ACO has been inspired by the behavior of real ant colonies, in particular,

by their foraging behavior. One of its main ideas is the indirect communication among the individuals of a colony of agents, called (artificial) ants, based on an analogy with trails of a chemical substance, called pheromone, which real ants use for communication [46]. The first ACO algorithm, called Ant System (AS), was applied to the Traveling Salesman Problem (TSP) [23, 20].

ACO algorithms make use of simple agents called ants which iteratively construct candidate solutions to a combinatorial optimization problem. The ants' solution construction is guided by (artificial) pheromone trails and problem-dependent heuristic information. In principle, ACO algorithms can be applied to any combinatorial optimization problem by defining solution components which the ants use to iteratively construct candidate solutions and on which they may deposit pheromone. An individual ant constructs candidate solutions, then, after the solution construction is completed, the ants give feedback on the solutions they have constructed by depositing pheromone on solution components which they have used in their solution. Typically, solution components which are part of better solutions or are used by many ants will receive a higher amount of pheromone, and hence, will more likely be used by the ants in future iterations of the algorithm. To avoid the search getting stuck, typically before the pheromone trails get reinforced, all pheromone trails are decreased by a factor ϱ.

The ants' solutions are not guaranteed to be optimal with respect to local changes and hence may be further improved using local search methods; take as example a 3-opt local search algorithm, that proceeds by systematically testing whether the current tour can be improved by replacing at most three arcs. Based on this observation, the best performing ACO algorithms for many NP-hard static combinatorial problems are in fact hybrid algorithms combining probabilistic solution construction by a colony of ants with local search algorithms. In such hybrid algorithms, the ants can be seen as guiding the local search by constructing promising initial solutions, because ants preferably use solution components which, earlier in the search, have been contained in good locally optimal solutions.

Traveling Salesman Problem. The TSP can be represented by a complete graph $G = (N, A)$ with N being the set of nodes, also called cities, and A being the set of arcs fully connecting the nodes. Each arc $(i, j) \in A$ is assigned a value d_{ij} which represents the distance between cities i and j. The TSP then is the problem of finding a shortest closed tour visiting each of the $n = |N|$ nodes of G exactly once.

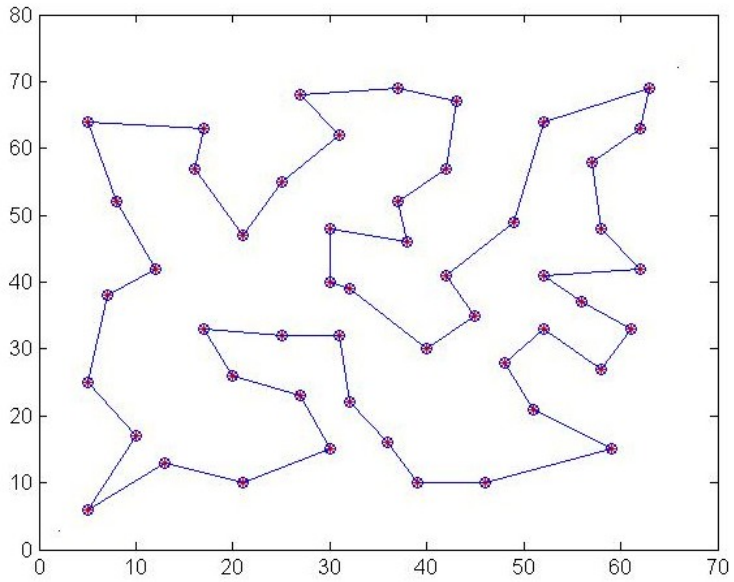


Figure 8.1: Best solution (Hamiltonian cycle with minimal total weight) of the Travel Salesman problem for *eil51*, a problem taken from the TSPLIB benchmark library [95]; there are 51 cities, each identified by a position $(x_i, y_i) \in \mathbb{R}^2$. The edge weight d_{ij} is given by the Euclidean distance between city i and city j.

Applying AS to the TSP. When applying AS to the TSP, arcs are used as solution components. A pheromone trail $\tau_{ij}(t)$, where t is the iteration counter, is associated with each arc $(i.j)$; these pheromone trails are modified during the run of the algorithm through pheromone trail evaporation and pheromone trail reinforcement by the ants. Initially, m ants are placed on m randomly chosen cities. Then, in each construction step, each ant moves, based on a probabilistic decision, to a city it has not yet visited. This probabilistic choice is biased by the pheromone trail $\tau_{ij}(t)$ and by a locally available heuristic information η_{ij}. The latter is a function of the arc length; AS and all other ACO algorithms for the TSP use $\eta_{ij} = 1/d_{ij}$. Ants prefer cities which are close and connected by arcs with a high pheromone trail and in AS an ant k currently located at city i chooses to go to city j with a probability:

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in \mathcal{N}_i^k} [\tau_{il}(t)]^\alpha [\eta_{il}]^\beta} \text{ if } j \in \mathcal{N}_i^k, \tag{8.1}$$

where α and β are two parameters which determine the relative importance of the pheromone trail and the heuristic information, and $\mathcal{N}_i^k$ is the feasible neighborhood of ant k, that is, the set of cities which ant k has not visited yet. Each ant k stores the cities visited in its current partial tour in a list, that is, each ant has a limited memory which is used to determine $\mathcal{N}_i^k$ in each construction step and thus to guarantee that only valid Hamiltonian cycles are generated. Additionally, it allows the ant to retrace its tour, once it is completed, so that it can deposit pheromone on the arcs it contains.

After all ants have completed the tour construction, the pheromone trails are updated. This is done first by lowering the pheromone trails by a constant factor (evaporation) and

then by allowing the ants to deposit pheromone on the arcs they have visited. In particular, the update follows this rule:

$$\tau_{ij}(t+1) = \varrho\tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t),$$

where the parameter ϱ (with $0 \leq \varrho < 1$) is the trail persistence (thus, $1-\varrho$ models the evaporation) and $\Delta\tau_{ij}^{k}(t)$ is the amount of pheromone ant k puts on the arcs it has used in its tour. The evaporation mechanism helps to avoid unlimited accumulation of the pheromone trails. While an arc is not chosen by the ants, its associated pheromone trail decreases exponentially; this enables the algorithm to "forget" bad choices over time. In AS, $\Delta\tau_{ij}^{k}(t)$ is defined as follows:

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} 1/L^{k}(t) & \text{if arc } (i,j) \text{ is used by ant } k \text{ in iteration } t, \\ 0 & \text{otherwise.} \end{cases}$$

where $L^{k}(t)$ is the tour length of the kth ant. The better the ant's tour is, the more pheromone is received by the arcs belonging to this tour. In general, arcs which are used by many ants and which are contained in shorter tours will receive more pheromone and therefore will more likely be chosen in future iterations of the algorithm.

8.1.2 $\mathcal{MAX}-\mathcal{MIN}$ Ant System.

Research on ACO has shown that improved performance may be obtained by a stronger exploitation of the best solutions found during the search. $\mathcal{MAX}-\mathcal{MIN}$ Ant System, which has been specifically developed to meet these requirements by T.Stutzle ([86, 87]), differs in three key aspects from AS.

1. To exploit the best solutions found during an iteration or during the run of the algorithm, after each iteration only one single ant adds pheromone. This ant may be the one which found the best solution in the current iteration (*iteration-best ant*) or the one which found the best solution from the beginning of the trial (*global-best ant*).

2. To avoid stagnation of the search the range of possible pheromone trails on each solution component is limited to an interval $[\tau_{min}, \tau_{max}]$.

3. Additionally, we deliberately initialize the pheromone trails to τ_{max}, achieving in this way a higher exploration of solutions at the start of the algorithm.

In $\mathcal{MM}$AS only one single ant is used to update the pheromone trails after each iteration. Consequently, the modified pheromone trail update rule is given by

$$\tau_{ij}(t+1) = \varrho\tau_{ij}(t) + \Delta\tau_{ij}^{\text{best}},$$

where $\Delta\tau_{ij}^{\text{best}} = 1/f\left(s^{\text{best}}\right)$ and $f\left(s^{\text{best}}\right)$ denotes the solution cost of either the iteration-best (s^{ib}) or the global-best solution (s^{gb}). The use of only one solution, either s^{ib} or s^{gb}, for the pheromone update is the most important means of search exploitation in $\mathcal{MM}$AS. When using only s^{gb}, the search may concentrate too fast around this solution and the exploration of possibly better ones is limited, with the consequent danger of getting trapped in poor quality

solutions. This danger is reduced when s^{ib} is chosen for the pheromone trail update since the iteration-best solutions may differ considerably from iteration to iteration and a larger number of solution components may receive occasional reinforcement.

Search stagnation may occur. This can happen if at each choice point, the pheromone trail is significantly higher for one choice than for all the others. In the TSP case, this means that for each city, one of the exiting arcs has a much higher pheromone level than the others. In this situation, due to the probabilistic choice governed by Eq. 8.1, an ant will prefer this solution component over all alternatives and further reinforcement will be given to the solution component in the pheromone trail update. In such a situation the ants construct the same solution over and over again and the exploration of the search space stops. By limiting the influence of the pheromone trails one can easily avoid the relative differences between the pheromone trails from becoming too extreme during the run of the algorithm. To achieve this goal, $\mathcal{MM}$AS imposes explicit limits τ_{min} and τ_{max} on the minimum and maximum pheromone trails such that for all pheromone trails $\tau_{ij}(t)$, $\tau_{min} \leq \tau_{ij}(t) \leq \tau_{max}$. Note that by enforcing $\tau_{min} > 0$ and if $\eta_{ij} < \infty$ for all solution components, the probability of choosing a specific solution component is never 0.

In $\mathcal{MM}$AS the pheromone trails are initialized in such a way that after the first iteration all pheromone trails correspond to τ_{max}. This type of trail initialization is chosen to increase the exploration of solutions during the first iterations of the algorithm. The experimental results confirm the conjecture that the larger exploration of the search space due to setting $\tau(1) = \tau_{max}$ improves $\mathcal{MM}$AS' performance.

The computational results given in the article show that $\mathcal{MM}$AS, in general, is able to find very high quality solutions for all instances from the TSPLIB benchmark library [95], that has been used in many other studies and partly stem from practical applications of the TSP; furthermore, for almost all instances $\mathcal{MM}$AS finds the optimal solution in at least one of the runs. This result shows the viability of the ant approach to generate very high quality solutions for the TSP. Note that the computational results with local search are also much better than those obtained without local search; additionally, the computation times with local search are much smaller.

8.1.3 Genetic Algorithms

In the 1950s and the 1960s several computer scientists independently studied evolutionary systems with the idea that evolution could be used as an optimization tool for engineering problems. The idea in all these systems was to evolve a population of candidate solutions to a given problem, using operators inspired by natural genetic variation and natural selection. In the 1960s, Rechenberg [79, 80] introduced "evolution strategies", a method he used to optimize real valued parameters for devices such as airfoils. Fogel, Owens, and Walsh [28] developed "evolutionary programming", a technique in which candidate solutions to given tasks were represented as finite state machines, which were evolved by randomly mutating their state transition diagrams and selecting the fittest.

Genetic algorithms (GAs) were invented by John Holland in the 1960s and were developed by Holland and his students and colleagues at the University of Michigan in the 1960s and the 1970s. In contrast with "evolution strategies" and "evolutionary programming", Holland's original goal was not to design algorithms to solve specific problems, but rather to formally study the phenomenon of adaptation as it occurs in nature and to develop ways in which the

mechanisms of natural adaptation might be imported into computer systems. Holland's 1975 book *Adaptation in Natural and Artificial Systems* [45] presented the genetic algorithm as an abstraction of biological evolution and gave a theoretical framework for adaptation under the GA. Holland's GA is a method for moving from one population of "chromosomes" (e.g., strings of ones and zeros, or "bits") to a new population by using a kind of "natural selection" together with the genetics inspired operators of crossover, mutation, and inversion. Each chromosome consists of "genes" (e.g., bits), each gene being an instance of a particular "allele" (e.g., 0 or 1). The selection operator chooses those chromosomes in the population that will be allowed to reproduce, and on average the fitter chromosomes produce more offspring than the less fit ones. Crossover exchanges subparts of two chromosomes, roughly mimicking biological recombination between two single chromosome (haploid) organisms; mutation randomly changes the allele values of some locations in the chromosome; and inversion reverses the order of a contiguous section of the chromosome, thus rearranging the order in which genes are arrayed.

Genetic algorithms have been used by many researchers as a tool for search and optimization. A given optimization task is encoded in such a way that instances such as a path in a weighted graph are understood as elements in a finite collection $\mathcal{C}$ of creatures (candidate solutions) in a model "world", and a fitness function $f : \mathcal{C} \longmapsto \mathbb{R}^+$ exists, which has to be maximized. Usually, the number of elements in $\mathcal{C}$ is very large prohibiting a complete search of $\mathcal{C}$. Genetic algorithms provide a probabilistic way to conduct a search in $\mathcal{C}$ for arbitrary f given a suitable encoding of creatures or instances into strings of symbols.

A genetic algorithm comprises three phases (operations): *mutation*, *crossover* and *fitness selection*. These are applied cyclically and iteratively to fixed-size, finite populations consisting of elements (chromosomes) of $\mathcal{C}$ until a saturation condition, or another boundary condition is satisfied. The model most commonly investigated is the genetic algorithm with a binary alphabet (where chromosomes take the form of bit strings, with each locus that has two possible alleles: 0 and 1), multiple bit mutation, one-point crossover, and proportional fitness selection.

Fitness selection models reproductive success of adapted organisms in their environment. This operator selects chromosomes in the population for reproduction. The fitter the chromosome, the more times it is likely to be selected to reproduce.

Crossover models the exchange of genetic information of creatures, inspired by exchange of genetic information in living organisms, e.g., during the process of sexual reproduction, and mimics biological recombination between two haploid organisms. This operator randomly chooses a locus and exchanges the subsequences before and after that locus between two chromosomes to create two offspring. For example, the strings 10000100 and 11111111 could be crossed over after the third locus in each to produce the two offspring 10011111 and 11100100.

Mutation models random change in the genetic information of creatures, and is inspired by random change of genetic information in living organisms, e.g., through the effects of radiation or chemical mismatch. This operator randomly flips some of the bits in a chromosome. For example, the string 00000100 might be mutated in its second position to yield 01000100. Mutation can occur at each bit position in a string with some probability, usually very small.

Below we present an example of GA, that will be used to test the procedure presented in the following sections.

- Take $4m$ individuals, binary strings of N elements; $\mathcal{C} = \{0,1\}^N$;

Repeat the following until the termination condition is obtained:

- Order the individuals accordingly to the value of the fitness function to be minimized:

$$f(x) = \left| \sum_{i=1}^{N} x_i - \frac{N-1}{2} \right|,$$

  so, for example, 11101 is listed above 00001 (see Fig. 8.2 to see f plotted versus the number of 1);

- Keep the first half individuals of the list, $2m$;

- Draw without replacement random pairs of individuals from the remaining $2m$;

- For each pair, two new individuals are produced by using single-point crossover; so, for example, from 00001 and 01101, by a single-point crossover in the third position, 01001 and 00101 are obtained;

- For each individual in the population, flip a component of the binary string chosen uniformly and independently, and replace the old chromosome only in the case that the new one increases the value of the fitness function; so, for example, a mutation of 00110 produces 10110 that has higher fitness;

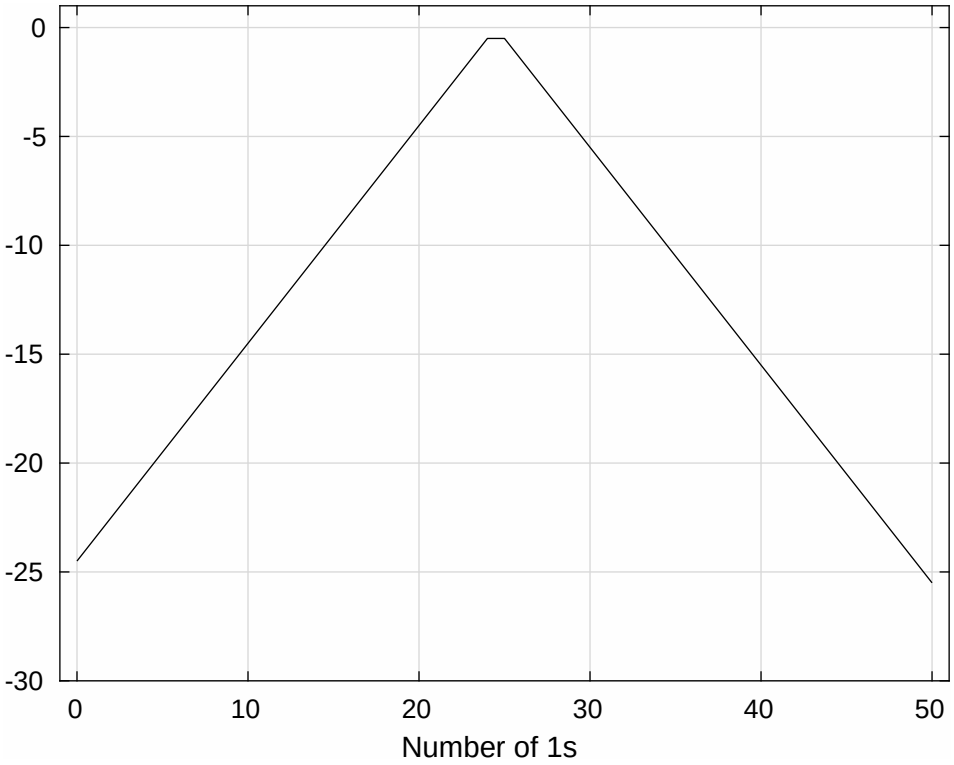


Figure 8.2: Plot of the considered pseudo-Boolean function value versus the number of 1s of the binary string.

A more detailed review of genetic algorithms can be found in the books [65, 82].

A natural issue for meta-heuristic algorithms (MHA) concerns their convergence [32, 37, 70, 38, 82]. Due to the stochastic nature of such algorithms, they have to be studied probabilistically; unfortunately, even when their convergence is theoretically guaranteed, it is often too slow to successfully use them in practice. One possible way to cope with this problem is the so called *restart* approach, which is used more generally for simulating rare events [30, 31, 64]. It consists of several independent executions of a given MHA: the executions are randomly initialized and the best solution, among those produced, is chosen. When implementing the restart on a non-parallel machine, the restart consists of periodic re-initialitations of the underlying MHA, the period T being called *restart time*. In the next chapter, we analyze in detail a new restart procedure, also using the algorithms presented above.

Chapter 9

A new restart procedure for combinatorial optimization

Despite the fact that the restart approach is widely used, very little work has been done to study it theoretically for combinatorial optimization [41, 66]. In [41], the restart is studied in its dynamic form instead of the static one considered here. Some results are provided for a specific evolutionary algorithm, i.e. the so called (1+1)EA, used to minimize three pseuso-Boolean functions. In [66] the fixed restart strategy is considered as done here. The first two moments of the random time T_R for the restart to find a solution (optimization time) are studied as a function of T. An equation for T is derived, whose solution minimizes the expected value of T_R. However, this equation involves the distribution of the optimization time of the underlying MHA, which is unknown.

In practice, the underlying MHA is very commonly restarted when there are negligible differences in the fitness of the best-so-far solutions at consecutive iterations during a certain time interval. This criterion may not be adequate when we want to really find the COP solution and we are not satisfied with suboptimal ones.

In this Chapter we propose a new iterative procedure to optimize the restart procedure for a combinatiorial optimization problem. Each iteration of the procedure consists of either adding new MHA executions or extending along time the existing ones. Along the iterations, the procedure uses an estimate of the MHA failure probability where the optimal solution is replaced by the *best so far* one. We recall that the failure probability of a stochastic algorithm $p(k)$ is the probability that the optimal solution has not been found up to iteration k. We prove that, with probability one, the restart time of the proposed procedure approaches, as the number of iterations diverges, the value that minimizes $\mathbb{E}[T_R]$.

We also show the results of the application of the proposed restart procedure to several TSP instances, whose solution is known, with hundreds or thousands of cities. As MHA we use different versions of the ACO algorithm Max-Min Ant System ($\mathcal{MM}$-AS). Based on a large number of experiments, we compare the results from the restart procedure with those from the MMAS. This is done by considering the failure probability of the two approaches for the same total computation cost. This comparison shows a significant gain when applying the proposed restart procedure. Similar results are obtained when applying the RP to the MMAS or to a GA for solving some instances of a pseudo-Boolean problem. The algorithms have been implemented in MATLAB or in C.

9.1 The procedure

In this section, we first give some mathematical details of the restart, that will be then used to define the RP. In the following part, we present the procedure in details, also providing a pseudocode. Following [12], the failure probability at iteration k of the restart is the probability that the optimization time of the restart T_R is larger than k:

$$\mathbb{P}(T_R > k) = p(T)^{\left\lfloor \frac{k-1}{T} \right\rfloor} p\left(k - \left\lfloor \frac{k-1}{T} \right\rfloor T\right), \tag{9.1}$$

where $p(t)$ is the failure probability of the underlying MHA, i.e. the probability that the optimal solution has not been found up to time t. We notice that the first factor in (9.1) is the probability that the optimal solution has not been found in each of the previous $\left\lfloor \frac{k-1}{T} \right\rfloor$ re-initializations. The second factor is the probability that the optimal solution has not been found up to iteration $k - \left\lfloor \frac{k-1}{T} \right\rfloor T$ of the current re-initialization. The restart failure probability is geometrically decreasing towards zero with the number $\left\lfloor \frac{k-1}{T} \right\rfloor$ of re-initializations, the base of such geometric sequence being $p(T)$. Therefore, a short restart time T may result in a high value of $p(t)$ and a slow convergence. On the contrary, if the restart time T is high, we may end up with a low number of re-initializations $\left\lfloor \frac{k-1}{T} \right\rfloor$ and high value of the restart failure probability. Then, a natural problem is to find an "optimal value" of T when using a finite amount of computation time.

The restart could be optimized by choosing a value for T that minimizes the expected value of the time T_R:

$$\mathbb{E}[T_R] = \sum_{k=1}^{\infty} \mathbb{P}(T_R > k). \tag{9.2}$$

In fact for any random variable not negative X it is possible to write

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > t)\, dt. \tag{9.3}$$

In our case, the random variable T_R is discrete and the integral in (9.3) is replaced by a series whose generic term is $\mathbb{P}(T_R > t)$.

We now derive an upper bound for the r.h.s. of (9.3):

$$\mathbb{E}[T_R] \leq \sum_{k=1}^{\infty} p(T)^{\left\lfloor \frac{k-1}{T} \right\rfloor} \leq \sum_{k=1}^{\infty} p(T)^{\frac{k-1}{T}-1} = \frac{1}{(1 - p(T)^{\frac{1}{T}})p(T)}. \tag{9.4}$$

By means of this bound, we can then optimize the RP by minimizing the function $g(x) := \left[(1 - p(x)^{\frac{1}{x}})p(x)\right]^{-1}$.

Whenever this function does not have a global minimum, there is no advantage to use the restart. In the other case, an optimal value for the restart time is provided by the first value t_m where the function g assumes its absolute minimum. However, this criterion cannot be applied in practice since the MHA failure probability is unknown.

The restart procedure (RP) starts by executing r_0 independent replications of the underlying MHA for a certain number of time steps T_0. Let us denote by $X_i(t)$ the solution produced by the replication i of the underlying algorithm at time t.

Let $Y_i(t)$ be the fitness function value of the best solution found by i-th replication until time t i.e. $Y_i(t) = \min\{f(X_i(s)), s = 1, ..., t\}$, where f is the function to minimize. Each $\{Y_i(t), t = 1, 2, \ldots\}$, $i = 1, 2, \ldots$ is an independent realization of the same process. Then, at the end of iteration k, based on the criterion described later in this section, the RP either increases the number of replications from r_k to r_{k+1} by executing $r_{k+1} - r_k$ replications of the underlying algorithm until time T_k, or it continues the execution of the existing r_k replications until time $T_{k+1} > T_k$. Therefore, the RP can be described by a sequence $\{Y_{A_k}, k = 0, 1, \ldots\}$ of nested finite matrices, extracted from the infinite matrix $\mathbf{Y}$:

$$\mathbf{Y} = \begin{pmatrix} Y_1(1) & Y_1(2) & \cdots & Y_1(t) & \cdots \\ Y_2(1) & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_i(1) & \cdots & \cdots & Y_i(t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix},$$

where $A_k := \{(i,t) : i = 1, \ldots, r_k \quad t = 1, \ldots, T_k\}$. The matrix Y_{A_k} corresponds to the first r_k rows and T_k columns of $\mathbf{Y}$. Let $\tilde{Y}_k$ denote the minimum value of this matrix at the end of iteration k: $\tilde{Y}_k = \min Y_{A_k} = \min\limits_{(i,t)\in A_k} Y_i(t)$. We estimate the failure probability sequence by means of the empirical frequency

$$\hat{p}_k(t) = \begin{cases} \dfrac{1}{r_k} \displaystyle\sum_{i=1}^{r_k} 1_{\{Y_i(t) > \tilde{Y}_k\}} & t = 1, \ldots, T_k, \\ 0 & \text{otherwise.} \end{cases}$$

Next, consider the function $g_k(t) = [(1 - \hat{p}_k(t)^{\frac{1}{t}})\hat{p}_k(t)]^{-1}$, $t = 1, \ldots, T_k$, and define $\hat{\sigma}_k$ the first time with a left and right increase of the function g_k (relative minimum). Let λ be a number in $(0, 1)$. If $\hat{\sigma}_k < \lambda \cdot T_k$, then the RP increases the number of replications by means of a certain rule $r_{k+1} := f_r(r_k)$. Otherwise, the RP increases the restart time according to $T_{k+1} := f_T(T_k)$. We assume that $\forall x$ we have $f_r(x) > x$ and $f_T(x) > x$. As a consequence, for any fixed $x > 0$, it holds $f_r^{(k)}(x), f_T^{(k)}(x) \to \infty$, k denoting the consecutive application of a function for k times. Therefore, the recursive formula for (r_k, T_k) is

$$(r_{k+1}, T_{k+1}) = \begin{cases} (f_r(r_k), T_k) & \text{if } \hat{\sigma}_k < \lambda \cdot T_k, \\ (r_k, f_T(T_k)) & \text{otherwise.} \end{cases}$$

Below there is the pseudocode for RP.

THE RP PSEUDOCODE:

$r = r_0$;

$T = T_0$;

for replication $i = 1, 2, \ldots, r$ **do**

- perform execution $\mathcal{A}_i$ of MHA until time T_0;
- save $\mathcal{A}_i(T_0)$;

end for
save Y_{A_0};
compute $\hat{\sigma}_0$ from Y_{A_0};
for iteration $k = 1, 2, \ldots$ **do**

- **if** $\hat{\sigma}_{k-1} \geq \lambda \cdot T_{k-1}$ **then**
  - $T_k = f_T(T_{k-1})$;
  - $r_k = r_{k-1}$;
  - **for** replication $i = 1, 2, \ldots, r_k$ **do**
    - continue the execution of $\mathcal{A}_i$ until T_k;
    - save $\mathcal{A}_i(T_k)$;
  - **end for**
- **else then**
  - $r_k = f_r(r_{k-1})$;
  - $T_k = T_{k-1}$;
  - **for** replication $i = r_{k-1} + 1, r_{k-1} + 2, \ldots, r_k$ **do**
    - execute $\mathcal{A}_i$ until T_k;
    - save $\mathcal{A}_i(T_k)$;
  - **end for**
- **end if**
- save Y_{A_k};
- compute $\hat{\sigma}_k$ from Y_{A_k};

end for

9.2 RP convergence

In this section, we describe some theoretical properties of the RP. The main result, i.e. Theorem 38, concerns with the RP convergence. Specifically, we prove that, woth probability one, the restart time of the RP approaches, as the iteration number diverges, the value t_m that minimizes $\mathbb{E}[T_R]$. This is done as follows. In Lemma 35, we prove that the number of replications r_k diverges, and that the RP eventually finds the optimal solution. This lemma is then used to prove the technical Lemma 36. Finally, this lemma and Theorem 37 are used to prove the main result.

We denote by f_m the value of the solution of the optimization problem. Moreover, we recall the functions $g(t) = [(1 - p(t)^{\frac{1}{t}})p(t)]^{-1}$, where $p(t)$ is the failure probability and $g_k(t) = [(1 - \hat{p}_k(t)^{\frac{1}{t}})\hat{p}_k(t)]^{-1}$, whose domain is $\{1, \ldots, T_k\}$.

In order to derive the following results, we assume that

1. $g(\cdot)$ admits absolute minimum at the time t_m of its first local one, and it is strictly decreasing for $t \leq t_m$,

2. $p(1) < 1$.

We notice that if the absolute minimum of $g(\cdot)$ is reached after the time t_m of its first local one, the RP restart time still converges to t_m, so providing only a suboptimal value for restart time. Point 2 in practice does not give limitations. In fact, we can always aggregate the first iteration of the algorithm with failure probablity less than one and those before it into a single one.

Remark 34. We notice that, by the assumptions on the functions f_r and f_T, and by the RP definition, the probability that both the sequences r_k and T_k are bounded is zero.

Lemma 35. *Let $p(t)$ be as above. Let (r_k, T_k) be the sequence of random variables which describes the RP. Then, it holds*

1. $\mathbb{P}(r_k \to \infty) = 1$,

2. $\mathbb{P}\left(\left\{\exists k : \quad \tilde{Y}_k = f_m\right\}\right) = 1$.

Proof. **1.** If $\mathbb{P}(r_k \to \infty) < 1$, then, with positive probability, the following three conditions hold for a certain positive integer r:

1. $r_k = r$ eventually;

2. T_k diverges (for Remark 34);

3. $\hat{\sigma}_k \geq \lambda T_k$ eventually (from 2.) and the definition of the RP).

Eventually, there are only two mutually exclusive possibilities: either the underlying r copies of the algorithm have all reached the optimum or only some $r' < r$ of them will have experienced it. In both cases, it follows that $\hat{p}_h(t)$ as well as $g_h(t)$ will not change for h large enough. Hence, eventually $\hat{\sigma}_h$ does not change as h increases, which is a contradiction with 3. Therefore $\mathbb{P}(r_k \to \infty) = 1$.

2. Since $r_k = r$ eventually, and since $p(1) < 1$, with probability one, there exists i such that $Y_i(1) = f_m$; for all k so large that $r_k \geq i$ it will be $\tilde{Y}_k = f_m$. This proves the point. □

Lemma 36. *For each $t \in \mathbb{N}$, it holds*

$$\mathbb{P}\left(\left\{\sup_k T_k < t\right\} \cup \left\{\sup_k T_k \geq t\,,\ \lim_{k\to\infty} \hat{p}_k(t) = p(t)\right\}\right) = 1 \tag{9.5}$$

Proof. Let us consider the case when the event $E_t := \{\sup_k T_k \geq t\}$ happens, then we can eventually compute $\hat{p}_k(s)$, for $s = 1, \ldots, t$. By point 1 of Lemma 35 and the strong law of large numbers, we get

$$\mathbb{P}\left(E_k\,,\ \lim_{k\to\infty} \frac{1}{r_k} \sum_{i=1}^{r_k} 1_{\{Y_i(t) > f_m\}} = p(t)\right) = \mathbb{P}(E_k)\,. \tag{9.6}$$

Hence, using point 2 of Lemma 35, we obtain that

$$\mathbb{P}\left(E_k\,,\ \lim_{k\to\infty}\frac{1}{r_k}\sum_{i=1}^{r_k}1_{\{Y_i(t)>f_m\}}=p(t)\right)$$

is equal to

$$\mathbb{P}\left(E_k\,,\ \lim_{k\to\infty}\frac{1}{r_k}\sum_{i=1}^{r_k}1_{\{Y_i(t)>\tilde{Y}_k\}}=p(t)\right).$$

Since, by definition

$$\hat{p}_k(t)=\frac{1}{r_k}\sum_{i=1}^{r_k}1_{\{Y_i(t)>\tilde{Y}_k\}},$$

we have

$$\mathbb{P}\left(E_k\,,\ \lim_{k\to\infty}\hat{p}_k(t)=p(t)\right)=\mathbb{P}\left(E_k\right),$$

from which the thesis follows. □

Theorem 37. *For the RP it holds*

$$\mathbb{P}\left(\sup_k T_k>\frac{t_m}{\lambda}\right)=1.$$

Proof. Let us assume that the thesis is not true. Then, there exists an integer number M such that $M\le\frac{t_m}{\lambda}$ and $\mathbb{P}\left(\{\sup_k T_k=M\}\right)>0$. On the event $\{\sup_k T_k=M\}$, by both Lemma 36 and the continuous mapping, we have the convergence $g_k(t)\to g(t)$, for any $t\le M$. This means that, for any $\varepsilon>0$ there is a positive probability that $\bigcap_{t=1}^{M}\{|g_k(t)-g(t)|<\varepsilon\}$, when k is large enough. Let us define $\tilde{M}:=\min(M,t_m)$. We then have $g_k(\tilde{M})<g(\tilde{M})+\varepsilon$ and $g_k(t)>g(t)-\varepsilon$ for any $1\le t<\tilde{M}$. Subtracting the first inequality from the last one, we obtain

$$g_k(t)-g_k(\tilde{M})>g(t)-g(\tilde{M})-2\varepsilon.$$

Since g is strictly decreasing until t_m, the r.h.s. of the last inequality is strictly larger than $g(\tilde{M}-1)-g(\tilde{M})-2\varepsilon$. By taking ε sufficiently small, we get $g_k(t)-g_k(\tilde{M})>0$ for any $t<\tilde{M}$. Hence, with a positive probability, we get eventually $\hat{\sigma}_k\ge\tilde{M}$.

If $M\le t_m$, then $\tilde{M}=M$ and with positive probability eventually we have $\hat{\sigma}_k\ge M$. Since $\hat{\sigma}_k\le T_k\le\sup_k T_k=M$, we have eventually $\hat{\sigma}_k=T_k=M$. For one of such k, it holds $\frac{\hat{\sigma}_k}{T_k}=1>\lambda$, so that, by the definition of the RP, at the following iteration with positive probability we have $T_{k+1}>T_k=M=\sup_k T_k$, which is impossible.

In the other case, $t_m<M\le\frac{t_m}{\lambda}$, we have $\tilde{M}=t_m$, and there is a positive probability that $t_m=\tilde{M}\le\hat{\sigma}_k\le T_k\le\sup_k T_k=M$ for k large enough; for any of these values of k, we get $\frac{\hat{\sigma}_k}{T_k}\ge\frac{t_m}{T_k}\ge\frac{t_m}{M}\ge\lambda$. As a consequence, with positive probability eventually we have $T_{k+1}=f_T(T_k)>T_k$, which is a contraddiction with $\sup_k T_k=M$. □

Theorem 38. *If we define* $T:=\left\lceil\frac{t_m}{\lambda}\right\rceil$, *it holds*

1. $\mathbb{P}\left(\bigcap_{t=1}^{T} \lim_{k\to\infty} \hat{p}_k(t) = p(t)\right) = 1,$

2. $\mathbb{P}\left(\lim_{k\to\infty} \hat{\sigma}_k = t_m\right) = 1.$

Proof. **1**. By Theorem 37, for any $t = 1, 2, \ldots, T$, with probability one we can eventually compute $\hat{p}_k(t)$. Hence, by the two statements of Lemma 35 and the strong law of large numbers, we get

$$\mathbb{P}\left(\bigcap_{t=1}^{T} \lim_{k\to\infty} \hat{p}_k(t) = p(t)\right) = 1,$$

that completes the proof of this point.

2. From 1 and the continuous mapping, with probability one, it holds

$$\lim_{k\to\infty} g_k(t) = g(t),$$

for $t = 1, \ldots, T$, with $T > t_m$. Therefore, the sequence $\hat{\sigma}_k$ converges to t_m. □

Remark 39. The efficiency of the RP depends on the expected value of the ratio $\sup_k T_k/t_m$. Although we do not have derived upper-bounds for this ratio, in all applications we performed, it remains sufficiently close to one.

9.3 Numerical Results

Below, we describe some results of the application of the RP to solve different instances of the TSP studied in [86] and two istances of a pseudo-Boolean problem. The underlying algorithm used here in the RP is mainly the ACO proposed in [86], known as MMAS; for the TSP instances, it is combined with different local search procedures [86]. In addition, for the pseudo-Boolean problem, we also use a GA with a population of 20 individuals. At each GA iteration, first the individuals are ordered accordingly to the value of the fitness function. Then, the first half individuals of the list are kept. The remaining individuals are replaced by new ones, obtained from the first group in the following way: we first draw without replacement random pairs of indivduals; for each pair, two new individuals are produced by using single-point crossover. The new population is finally obtained by flipping a component of the binary string chosen uniformly and independently for each individual, in the case that this decreases the value of the fitness function.

The RP setting is as follows: $r_{k+1} = f_r(r_k) := c_1 \cdot r_k$ and $T_{k+1} = f_T(T_k) := c_2 \cdot T_k$ $ where $c_1 = 1.2, c_2 = 1.1$. The initial values for r_0 and T_0 are 20 and 100, respectively. Finally, we set $\lambda = \frac{4}{5}$.

For both the TSP instances and the pseudo-Boolean problem considered here, the optimal solution is known. This information can be used to estimate the failure probability of the RP and of the underlying algorithm. However, obviously this information cannot be used when applying the RP.

In order to compare the results from the two algorithms with the same computational effort, we consider for the RP a pseudo-time t, defined as follows: for the initial RP iteration, the first T_0 instants of the pseudo-time correspond to the first T_0 iterations of the first replication

of the underlying algorithm; the following T_0 pseudo-time instants correspond to the first T_0 iterations of the second replication and so on until replication r_0. At the end of the k-th RP iteration, we have produced r_k executions (replications) for T_k times and the final pseudo-time instant is $t = r_k \cdot T_k$. At the $(k+1)$-th iteration, we have a certain (r_{k+1}, T_{k+1}), with either $r_{k+1} > r_k$ and $T_{k+1} = T_k$ or $r_{k+1} = r_k$ and $T_{k+1} > T_k$. In the first case, the pseudo-time instant $t = T_k \cdot r_k + 1$ corresponds to the first iteration time of the $r_k + 1$ replication and it is increased until the end of that replication. We proceed in the same way until the end of r_{k+1} replication. In the second case, the pseudo-time instant $t = T_k \cdot r_k + 1$ corresponds to the iteration time $T_k + 1$ of the first replication and is then increased until the iteration time T_{k+1} of that replication. Then, the same procedure is applied for the remaining replications based on their number.

We denote by $\tilde{Y}(t)$ $(t = 1, 2, \dots)$ the process describing the best so far solution of the RP (MMAS or GA) corresponding to the pseudo-time (time) instant t. Hence, based on a set of m independent replications of the RP, we can estimate the failure probability $p_{RP}(t)$ that the optimal solution has not been found up to pseudo-time t by using the classical estimator

$$\hat{p}_{RP}(t) = \frac{1}{m} \sum_{i=1}^{m} 1_{\{\tilde{Y}_i(t) \neq f_m\}}, \tag{9.7}$$

and analogously with $\hat{p}(t)$ for the MMAS or GA. By the law of large numbers this estimator converges to the failure probability $p_{RP}(t)$ (to $p(t)$ for the MMAS or GA).

We start with the example where we want to minimize the following pseudo-Boolean function

$$f(x) = -\left| \sum_{i=1}^{N} x_i - \frac{N-1}{2} \right|, \tag{9.8}$$

with respect to all binary strings of length N. In Fig. 9.1, this function is plotted versus the number of 1s in the case of $N = 50$ considered now.

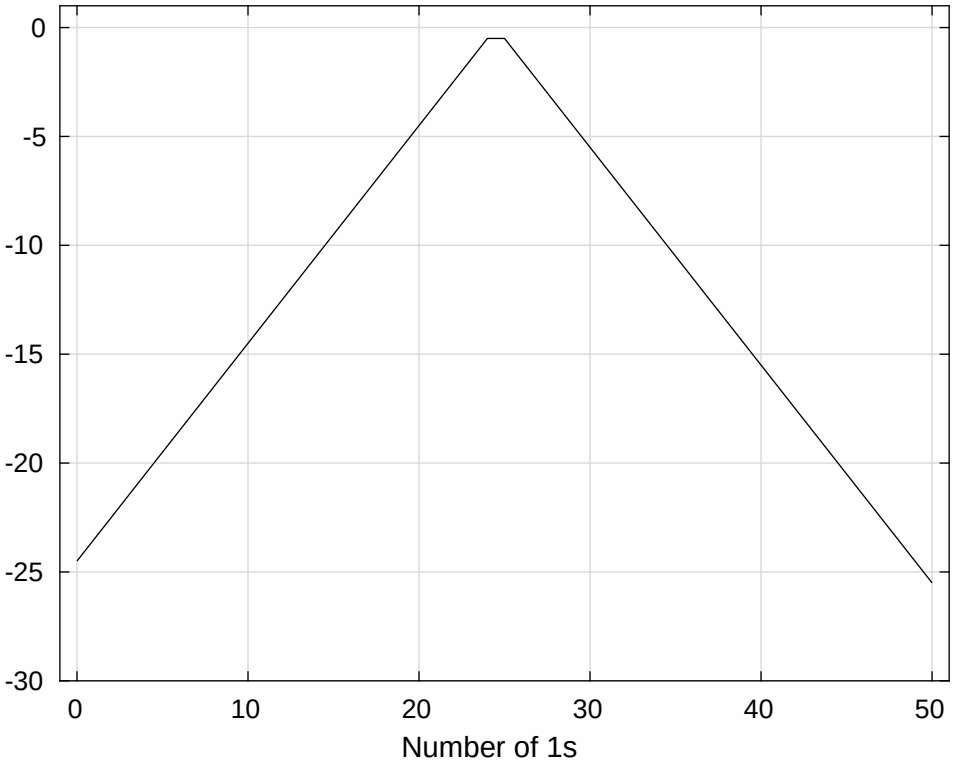


Figure 9.1: Plot of the considered pseudo-Boolean function value versus the number of 1s of the binary string.

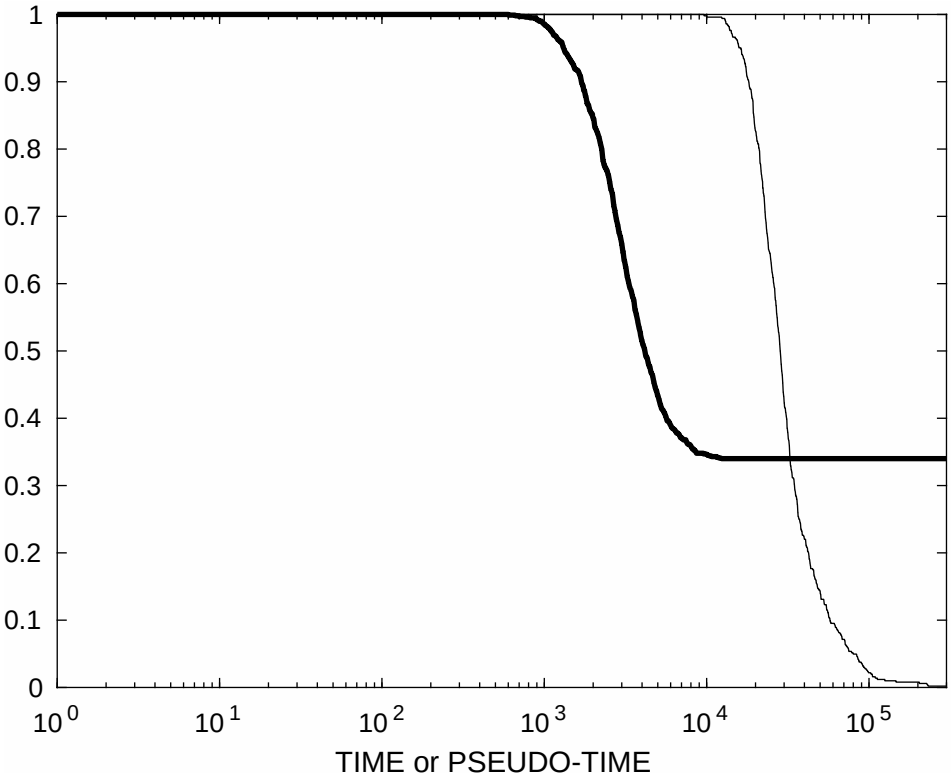


Figure 9.2: Pseudo-Boolean problem. The estimated failure probability as function of time or pseudo-time for the standard MMAS (thick line) and the RP (thin line).The time axis is in logarithmic scale. The f.p. curves of both the RP and the underlying algorithm are computed by the estimator in (9.7) based on 500 and 1000 replications, respectively.

This function has two local minima but only one of them is global. The first base algorithm considered is the MMAS, for which the pheromone bounds τ_{min} and $\tau_{\max}$ ensure that at any time, there is a positive probability to visit each configuration, e.g. the global minimum. Therefore, with probability one this algorithm will find the solution. However, if it reaches a configuration with few 1s, it takes in average an enormous amount of time, not available in practice, to move towards the global minimum. Therefore, we expect that in this case the restart will be successful.

In Fig. 9.2, we show the estimated failure probability (f.p.) $\hat{p}(t)$ for the MMAS algorithm to minimize the pseudo-boolean function of Fig. 9.1 (thick line). In the same figure, the estimated f.p. $\hat{p}_{RP}(t)$ of the RP is plotted versus the pseudo-time (thin line). We notice that there is a clear advantage to use the RP when compared to the standard MMAS. For this problem, we also use the GA described before. In this case we have $N = 300$ (boolean300). The curves obtained are very similar to those in Fig. 9.1. The f.p. estimated values are shown in Table 9.1.

We consider now an instance of the TSP with 532 cities (att532) solved by MMAS with the same setting as in [86]. After five hundreds of thousands of iterations, the underlying algorithm has an estimated f.p. of 0.38 ca. Instead, at the same value of the pseudo-time, the RP has a significantly lower f.p. (0.004 ca), as clearly shown in Fig.9.3. We remark that, until

the value 3900 ca for the time or pseudo-time, the f.p. of the underlying algorithm is lower than the one of RP. This is due to the fact that the RP is still learning the optimal value of the restart time. After that, the trend is inverted: the RP overcomes the MMAS and gains two orders of magnitude for very large values of the pseudo-time.

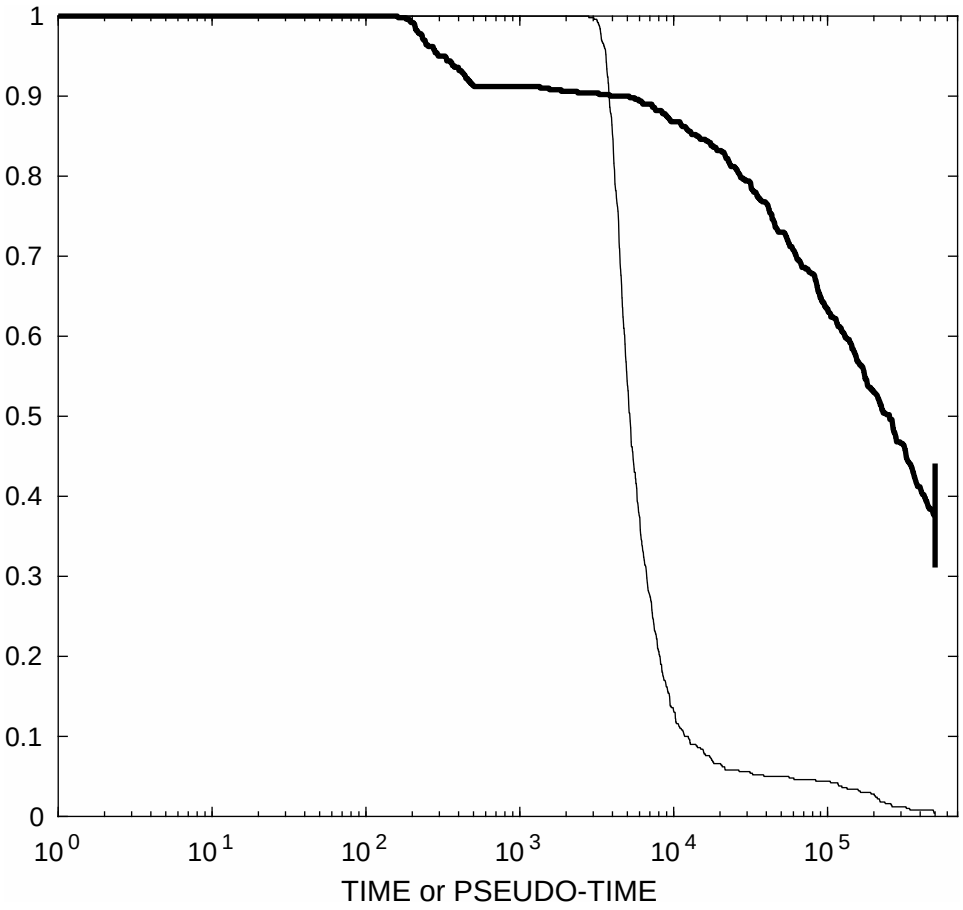


Figure 9.3: The TSP instance with 532 cities (att532). The estimated failure probability as function of time or pseudo-time for the standard MMAS (thick line) and the RP (thin line). The time axis is in logarithmic scale. The f.p. curves of both the RP and the underlying algorithm are computed by the estimator in (9.7) based both on 500 replications. The vertical segment shows the 99% level confidence interval.

We notice that the value $\hat{\sigma}_k$ approaches the optimal restart time t_m. In fact, as an example, in Fig. 9.4, we show the denominator of the function $g_k(t)$ at the end of a single RP execution. A global maximum appears at approximately the value of 430, the difference with the value of t_m, computed from the estimate $\hat{p}(t)$, being less than 1%.

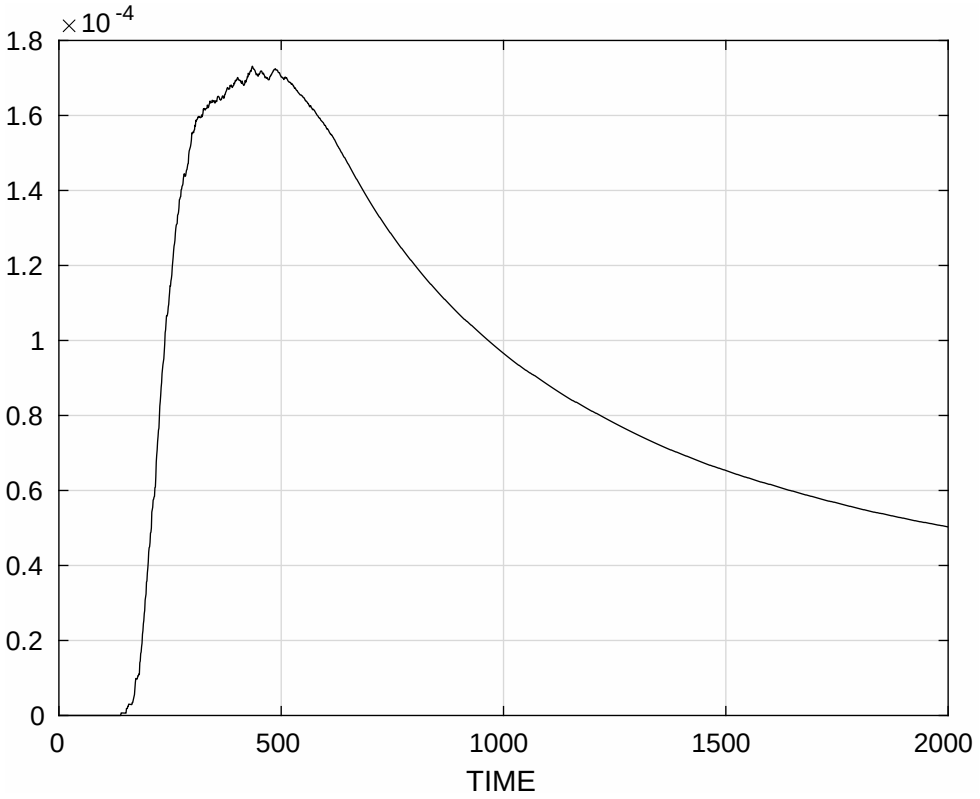


Figure 9.4: The TSP instance with 532 cities (att532). The denominator of the function $g_k(t)$ at the end of a single RP replication.

Finally in Fig. 9.5, we compare the f.p. curve for the RP with the one obtained applying the restart with the estimated optimal restart time. We remark that to use this approach in practice would require much longer computation than to execute the RP. In fact, we need to provide first an estimate of t_m by means of a sufficiently large sample of independent runs of the base algorithm. We notice that the RP curve starts to decrease significantly after the other one. This is due to the fact that the RP is still searching for the optimal value of the restart, whereas it is set from the beginning in the other (ideal) case. At about pseudo-time 7000, the two f.p.s become almost equal. After that, the f.p. of the MMAS goes to zero faster, even if the difference between the two f.p.s remains less than 0.05 ca. Finally, at pseudo-time $5 \cdot 10^5$, the estimated f.p. value of the RP is $4 \cdot 10^{-3}$, comparable to the estimated f.p. value of the standard restart at same time.

We notice that curves similar to those as in Fig.9.3, 9.4 and 9.5 were obtained for all the other TSP instances considered. The relative results are shown in Table 9.1.

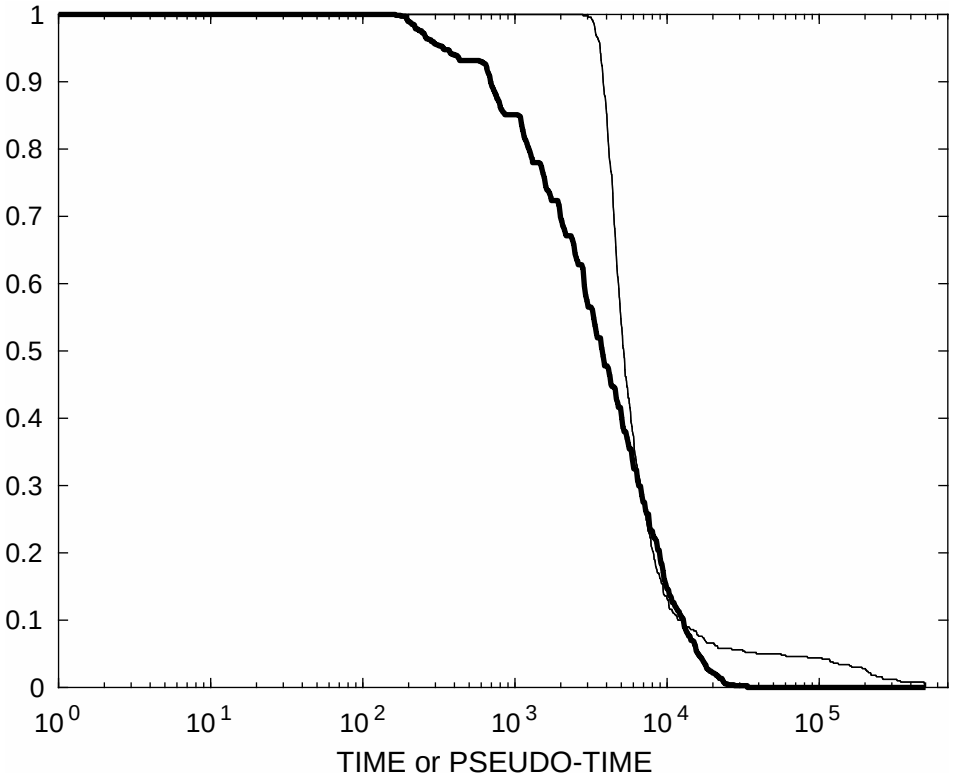


Figure 9.5: The TSP instance with 532 cities (att532). Comparison between the failure probability curve of the RP that appears in Fig. 9.3 (thin line) and the one obtained applying the restart periodically with the optimal restart time (thick line). The f.p. curves of both the RP and the underlying algorithm are computed by the estimator in (9.7) based on 500 replications.

Istance	**Base MHA**	$\mathbf{T_c}$	**Base MHA f.p..**	**RP f.p.**
boolean50	MMAS	300000	0.34	$2.1 \cdot 10^{-3}$
boolean300	GA	10000	0.43	0
pcb442	MMAS-3opt	100000	0.22	$4.0 \cdot 10^{-3}$
att532	MMAS-3opt	500000	0.38	$4.0 \cdot 10^{-3}$
lin318	MMAS-2.5opt	30000	0.44	0
d1291	MMAS-3opt	700000	0.57	$2.0 \cdot 10^{-3}$
d198	MMAS-2.5opt	100000	0.67	0

Table 9.1: Results of the application of the RP and the base MHA to TSP istances with known optimal solution and to istances of a preudo-boolean problem. The estimated f.p. values are computed at the time T_c reported in the third column (pseudo-time for the RP) based on at least 500 elements.

By looking at the results in Table 9.1, it is evident the advantage of using the RP instead of the underlying algorithm. In fact, for each case, the estimated f.p. value of the RP is several orders of magnitude lower than the one of the base MHA.

9.4 Conclusions

Given a combinatorial optimization problem, it is often needed to apply stochastic algorithms exploring the space using a general criterion independent of the problem. Unfortunately, usually there is a positive probability that the algorithm remains in a sub-optimal solution. This drawback can be coped by applying periodic algorithm re-initializations. This strategy is called restart. Although it is often applied in practice, there are few works studying it theoretically. In particular, there are no theoretical information to be used in practice to choose a convenient value for the restart time.

In this chapter, we propose a new procedure to optimize the restart (RP) and we study it theoretically. The iterative procedure starts by executing a certain number r_0 of independent replications of the underlying algorithm for a predefined time T_0. At the end of any iteration k of the RP, we have r_k independent replications each composed by T_k iterations of the underlying algorithm. We then compute the minimum value $\tilde{Y}_k$ of the objective function $f(\cdot)$ on these $r_k T_k$ points. Hence, for each time $t = 1, \ldots, T_k$, we compute an estimate $\hat{p}_k(t)$ of the failure probability $p(t)$, i.e. the probability that we have not yet reached the value $\tilde{Y}_k$. We now consider the function $g_k(t) = [(1 - \hat{p}_k(t)^{\frac{1}{t}})\hat{p}_k(t)]^{-1}$, that is the analogous of the function $g(t) = [(1 - p(t)^{\frac{1}{t}})p(t)]^{-1}$. We recall that the first time t_m at which the absolute minimum of $g(t)$ is reached corresponds to an "optimal value" of the restart time, that minimizes the expected time to find a solution. We then compute the position $\hat{\sigma}_k$ of the first minimum of $g_k(t)$. If $\hat{\sigma}_k$ is close to the end of the current execution time frame T_k of the underlying algorithm, T_k is increased; otherwise this is done for the number of replications r_k. This is controlled by the parameter $\lambda \in (0, 1)$.

The theory predicts that the RP eventually will find the optimal value of the restart time. In fact, the theorems prove that, with probability one, $\hat{p}_k(t)$, $g_k(t)$ and $\hat{\sigma}_k$ converge to $p(t)$, $g(t)$ and t_m, respectively.

In this work, we illustrate some results obtained by applying the RP to two versions of the MMAS ACO algorithm [86] for solving several TSP instances, whose solution is known, with hundreds or thousands of cities. The results obtained show that the estimated values of the failure probability of the RP are several orders of magnitude lower than those of the underlying algorithm, for equal computational cost. Therefore, given a certain computation resource, by applying the RP, we are far more confident that the result obtained is a solution of the COP instance analyzed. The procedure proposed could be improved preserving its performance and decreasing the computational cost. A possible way to do it is to increase the parameter λ along iterations. In fact, once we have a reasonably good estimate of $g(t)$, we would like to reduce the possibility that, by chance, we increase too much the time interval length. This can be done by increasing the value of λ.

Bibliography

[1] S.E. Ahnert. *Structural properties of genotype–phenotype maps.* J. R. Soc. Interface 14: 20170275, 2017.

[2] A.L. Amadori. *Obstacle problem for nonlinear integro-differential equations arising in option pricing,* Ricerche di Matematica, 56 (1) (pages 1-17), 2007.

[3] A. Amadori, M. Briani, R. Natalini. *A non-local rare mutations model for quasispecies and Prisoner's dilemma: numerical assessment of qualitative behaviour*, European Journal of Applied Mathematics, 2015.

[4] A. Amadori, A. Calzolari, R. Natalini, B. Torti. *Rare Mutations in Evolutionary Dynamics*, J.Differential Equations, 2015.

[5] A. L. Amadori, R. Natalini, D. Palmigiani. *A rare mutation model in a spatial heterogeneous environment.* Ecological Complexity, Volume 34 (pages 188-197), 2017.

[6] D.L. Applegate, R.M. Bixby, V. Chvatal, W.J. Cook. *The Traveling Salesman Problem.* Princeton University Press, 2006.

[7] K.B. Athreya, W. Kliemann, G. Koch. *On sequential construction of solutions of stochastic differential equations with jump terms.* Journal Systems & Control Letters 10 (2) (pages 141-146), Elsevier Science Publishers B. V. Amsterdam, 1988.

[8] A. Bazzani, M. Buiatti, P. Freguglia. *Metodi matematici per la teoria dell'evoluzione.* Springer-Verlag Italia, 2011.

[9] C.Blum, A.Roli. *Metaheuristics in combinatorial optimization: overview and conceptual comparison.* ACM Computing Surveys 3 (pages 268-308), 2003.

[10] I. Bomze, R. Buerger. *Stability by mutation in evolutionary games.* Games Econ. Behav. 11, (pages 146–172). 1995.

[11] G.E.P. Box, M.E. Muller. *A Note on the Generation of Random Normal Deviates.* The Annals of Mathematical Statistics 2, Vol. 29 (pages 610-611), 1958.

[12] L. Carvelli, G. Sebastiani. *Some issues of ACO algorithm convergence.* Ant Colony Optimization - Methods and Applications (pages 39-52). InTech, 2011.

[13] S. Ciliberti, O.C. Martin, A. Wagner. *Innovation and robustness in complex regulatory gene networks.* Proc. Natl Acad. Sci. USA 104 (pages 13591 - 13596), 2007.

[14] N. Champagnat, R. Ferriere, S. Meleard. *Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models.* Theoretical Population Biology, 69 (3), 2006.

[15] A. Croes. *A method for solving Traveling Salesman Problem.* Opns. Res. Volume 5 (pages 791-812), 1958.

[16] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes.* Springer, 2008.

[17] C. Darwin. *On the origin of species.* London, UK: John Murray, 1859.

[18] F. Darwin, *The Life and Letters of Charles Darwin.* London, UK: John Murray, 1887.

[19] C.M. Dobson, A. Sali, M. Karplus. *Protein folding: a perspective from theory and experiment.* Angew. Chem. Int. Ed. 37 (pages 868 - 893), 1998.

[20] M. Dorigo. *Optimization, Learning and Natural Algorithms*, Politecnico di Milano, PhD thesis, 1992.

[21] M. Dorigo, G. Di Caro. *The Ant Colony Optimization metaheuristic.* New Ideas in Optimization, 1999.

[22] M. Dorigo, V. Maniezzo, A. Colorni. *Ant System: Optimization by a colony of cooperating ants.* IEEE Trans. on System, Man, and Cybernetics, Volume 26, 1996.

[23] M. Dorigo, T. Stutzle. *Ant Colony Optimization.* MIT Press, 2004.

[24] M. Eigen e P. Schuster. *The hypercycle: a principle of natural self-organization.* Springer-Verlag, 1979.

[25] D. Eppstein. *Growth and Decay in Life-Like Cellular Automata.* Game of Life Cellular Automata. Springer-Verlag London Limited, 2010.

[26] L. C. Evans. *An introduction to stochastic differential equations.* American Mathematical Society, 2010.

[27] R.A. Fisher. *XV.—The correlation between relatives on the supposition of Mendelian inheritance.* Trans. R. Soc. Edinb. 52 (pages 399 - 433), 1919.

[28] D.B. Fogel, W. Atmar. *Proceedings of the Second on Evolutionary Programming.* Evolutionary Programming Society, 1993.

[29] M.Gardner. *The fantastic combinations of John Conway's new solitaire game "Life"*, Scientific American, 223 (pages 120 - 123), 1970.

[30] M.J.J. Garvels, D.P. Kroese. *A comparison of RESTART implementations.* Simulation Conference Proceedings (pages 601-608). IEEE Computer Society Press, 1998

[31] M.J.J. Garvels, D.P. Kroese. *On the entrance distribution in RESTART simulation.* Proceedings of the Rare Event Simulation (RESIM '99) Workshop (pages 65-88). University of Twente, 1999.

[32] S. Geman, D. Geman. *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 6 (pages 721-741),1984.

[33] F. Glover. *Tabu search-Part 1.* ORSA Journal of Computing, 3, Volume 1 (pages 190-206), 1989

[34] D.E. Goldberg, B. Korb, K. Deb. *Messy Genetic Algorithms: Motivation, analysis, and first results.* Complex Systems, Volume 3 (pages 493-530), 1989.

[35] S.J. Gould. *The Structure of Evolutionary Theory.* Belknap Press, 2002.

[36] S.F. Greenbury, S.E. Ahnert. *The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype– phenotype maps.* J. R. Soc. Interface 12: 20150724, 2015.

[37] W.J. Gutjhar. *A generalized convergence result for the graph-based ant system.* Probability in the Engineering and Informational Sciences, Volume 17 (pages 545-569), 2003

[38] W.J. Gutjhar, G. Sebastiani. *Runtime analysis of ant colony optimization with best-so-far reinforcement.* Methodology and Computing in Applied Probability, Volume 10 (pages 409-433), 2008.

[39] J.B.S. Haldane. *The causes of evolution.* Princeton, NJ: Princeton University Press, 1990.

[40] R.W. Hamming. *Error-detecting and error-correcting codes.* Bell System Technical Journal 29, Vol 2 (pages 147-160), 1950.

[41] T. Hansen. *On the analysis of dynamic restart strategies for evolutionary algorithms.* Proceedings of the 7th International Conference on Parallel Problem Solving from Nature (pages 33-43). Springer-Verlag, London, 2002.

[42] G.H. Hardy, *Mendelian proportions in a mixed population*, in Science., vol. 28, 1908.

[43] D. J. Higham. *An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations.* SIAM Review, Vol. 43, No. 3, pp. 525-546, 2001.

[44] J. Hofbauer e K. Sigmund. *Evolutionary games and population dynamics.* Cambridge University Press, 1998.

[45] J.H. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, 1975.

[46] B. Holldobler, E.O. Wilson. *The Ants.* Harvard University Press, 1990.

[47] H.H. Hoos, T. Stutzle. *Stochastic local search foundations and applications.* Morgan Kaufmann Publishers, 2004.

[48] W. Huang, B. Haubold, C. Hauert, A. Traulsen. *Emergence of stable polymorphism driven by evolutionary games between mutants.* Nature Communications 3, 919, 2012.

[49] N. Ikeda, S. Watanabe. *Stochastic Differential Equations and diffusion processes.* Kodansha Ltd., 1981.

[50] Ken-Iti Sato. *Levy Processes and Infinitely Divisible Distributions.* Cambridge: Cambridge University Press, 1999.

[51] M. Kimura. *Evolutionary rate at the molecular level.* Nature 217 (pages 624 - 626), 1968.

[52] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. *Optimization by simulated annealing.* Science, Volume 220 (pages 671-680), 1983.

[53] T. Kuusi, G. Mingione, Y. Sire. *Nonlocal Equations with Measure Data.* Communications in Mathematical Physics, 337 (3), 2015.

[54] A.J. Lotka. *Analytical Note on Certain Rhythmic Relations in Organic Systems.* Proc. Natl. Acad. Sci. U.S.A. 6 (pages 410-415), 1920.

[55] T. R. Malthus. *An Essay on the Principle of Population.* J. Johnson, London, 1798.

[56] S. Martello, P. Toth. *Knapsack Problems: Algorithms and Computer Implementation.* John Wiley and Sons, 1990.

[57] J. Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

[58] J. Maynard Smith. *Game Theory and The Evolution of Fighting.* On Evolution. Edinburgh University Press, 1972.

[59] J. Maynard Smith. *Natural selection and the concept of a protein space.* Nature 225 (pages 563 - 564), 1970.

[60] J. Maynard Smith, G.R. Price. *The logic of animal conflict.* Nature, 246 (pages 15 -18), 1973.

[61] E. Mayr. *Animal Species and Evolution.* Cambridge: Belknap P. of Harvard U.P., 1963.

[62] E. Mayr. *The Growth of Biological Thought.* Cambridge (Mass.): Belknap P. of Harvard U.P, 1982.

[63] G. Mendel. *Verhandlungen des naturforschenden.* Vereines in Brunn 4 (pages 3 - 47), 1866.

[64] A. Misevicius. *Restart-based genetic algorithm for the quadratic assignment problem.* Research and Development in Intelligent Systems XXV (pages 91-104). Springer-Verlag, 2009.

[65] M. Mitchell. *An introduction to genetic algorithms.* Cambridge, MA: MIT Press, 1996.

[66] A.P.A. Van Moorsel, K. Wolter. *Analysis and algorithms for restart.* Proceedings of the 1st International Conference on the Quantitative Evaluation of Systems (QEST) (pages 195-204), 2004.

[67] O. Morgenstern, J. Neumann, *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[68] P. Morters e Y. Peres. *Brownian Motion.* Cambridge Series in Statistical and Probabilistic Mathematics, 2010.

[69] J. Nash, *Non-Cooperative Games*. PhD thesis. Princeton University, 1950.

[70] F. Neumann, C. Witt. *Runtime Analysis of a simple ant colony optimization algorithm.* Algorithmica, Volume 45 (pages 243-255), 2007.

[71] J. Neumann, *Zur Theorie der Gesellschaftsspiele*. Mathematische Annalen, 100, Vol. 1 (pages 295-320), 1928.

[72] J.R. Norris. *Markov Chains.* Cambridge University Press, 1997.

[73] M. A. Nowak. *Evolutionary dynamics. Exploring the equations of life.* Cambridge, MA: The Belknap Press of Harvard University Press, 2006.

[74] D. Nualart. *The Malliavin Calculus and Related Topics.* New York, Springer Science+Business Media, 2006.

[75] B. Oksendal. *Stochastic Differential Equations.* Springer-Verlag, 2000.

[76] G. A. Pavliotis. *Stochastic Processes and Applications.* Springer. Texts in Applied Mathematics, 2014.

[77] J.L. Payne, A. Wagner. *Constraint and contingency in multifunctional gene regulatory circuits.* PLOS Comput. Biol. 9, e1003071, 2013.

[78] A. Quarteroni, R. Sacco, F. Saleri. *Numerical Mathematics.* Springer Berlin Heidelberg, 2007.

[79] I. Rechenberg. *Cybernetic Solution Path of an Experimental Problem.* Ministry of Aviation, Royal Aircraft Establishment (U.K.), 1965.

[80] I. Rechenberg. *Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution.* Frommann−Holzboog (Stuttgart), 1973.

[81] F.B. Salisbury. *Natural selection and the complexity of the gene.* Nature 224 (pages 342 - 343), 1969.

[82] L.M. Schmitt. *Fundamental Study - Theory of genetic algorithms.* Theoretical Computer Science 259 (pages 1-61). 2001.

[83] G. Sebastiani, D. Palmigiani. *Optimal Algorithm Re-initialization for Combinatorial Optimization*, Journal of Numerical Mathematics and Stochastics. 2019, *accepted.*

[84] A.V. Skorokhod. *Asymptotic methods in the theory of stochastic differential equations.* Translations of Mathematical Monographs 78. American Mathematical Society. 1989.

[85] P. Stadler and P. Schuster. *Mutation in autocatalytic reaction networks.* Journal of mathematical biology, 30 (6) (pages 597-632), 1992.

[86] T. Stutzle, H.H. Hoos. *Max-min ant system.* Future Generation Computer Systems, Volume 16 (pages 889-914), 2000.

[87] T. Stutzle H.H. Hoos. *The MAX-MIN Ant System and local search for the travelling salesman problem.* Proc.˜ICEC '97 Int. Conf. on Evolutionary Computation (pages 309-314), 1997.

[88] P.D. Taylor, L. Jonker. *Evolutionary stable strategies and game dynamics.* Math. Biosci. 40 (pages 145-156).

[89] A. Turing. *On computable numbers, with an application to the entscheidungsproblem.* Proc. London Math Soc., Volume 42, 1936.

[90] P.E. Turner, L. Chao. *Prisoner's dilemma in a RNA virus.* Nature, Vol. 398 (pages 441-443), 1999.

[91] P.E. Turner, L. Chao. *Escape from Prisoner's Dilemma in RNA Phage* Φ6. The American Naturalist, Vol. 161 (pages 497 - 505), 2003.

[92] P.F. Verhulst. *Recherches mathématiques sur la loi d'accroissement de la population.* Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles, 18 (pages 1-42), 1845.

[93] V. Volterra. *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi.* Mem. Acad. Lincei Roma. 2 (pages 31 - 113), 1926.

[94] S. Wright. *The roles of mutation, inbreeding, crossbreeding and selection in evolution.* In Proc. 6th Int. Congress on Genetics, Ithaca, NY, USA, vol. 1 (pages 356 - 366), 1932.

[95] https://wwwproxy.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/